

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2004, 14:44:25 ; Search time 0.787948 Seconds  
(without alignments)  
2731.623 Million cell updates/sec

Title: US-10-719-385-26  
Perfect score: 36  
Sequence: 1 AGGPGV 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues  
Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database : A\_Geneseq.23Sep04:.\*  
1: geneseq1980s:.\*  
2: geneseq1990s:.\*  
3: geneseq2000s:.\*  
4: geneseq2001s:.\*  
5: geneseq2002s:.\*  
6: geneseq2003s:.\*  
7: geneseq2003bs:.\*  
8: geneseq2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	100.0	6	AD071786	AD071786 BFLP0169
2	36	100.0	166	ABU70740	ABU70740 Human adi
3	36	100.0	407	ADJ69011	ADJ69011 Human hea
4	36	100.0	415	AAU27706	AAU27706 Human ful
5	36	100.0	415	ADJ58205	ADJ58205 Human Cez
6	36	100.0	479	ABB97204	ABB97204 Novel hum
7	36	100.0	843	AD126089	AD126089 Human pro
8	36	100.0	843	AD126087	AD126087 Human pro
9	36	100.0	858	ABR47408	ABR47408 Breast ca
10	36	100.0	858	ADG37584	ADG37584 Human nuc
11	36	100.0	1745	ADP76869	ADP76869 Novel hum
12	36	100.0	1745	AD071781	AD071781 BFLP0169
13	36	100.0	1767	AD071782	AD071782 BFLP0169
14	36	100.0	1883	ADG08006	ADG08006 Novel hum
15	35	97.2	142	AAG79569	AAG79569 ACRI C-te
16	35	97.2	1036	ABR82256	ABR82256 Human act
17	35	97.2	1068	ADG08157	ADG08157 Novel pro
18	33	91.7	642	ABB60188	ABB60188 Drosophila
19	32	88.9	18	AAW66643	AAW66643 HSV-2 gly
20	32	88.9	20	AAK45552	AAK45552 CRY J I P
21	32	88.9	20	AAK45551	AAK45551 CRY J I P
22	32	88.9	20	AAK82501	AAK82501 CRY J I J
23	32	88.9	20	AAK82500	AAK82500 CRY J I J
24	32	88.9	20	ADD93845	ADD93845 Japanese
25	32	88.9	20	ADD93844	ADD93844 Japanese

26	32	88.9	30	2	AAK45579	AAK45579 CRY J I P
27	32	88.9	30	2	AAK44683	AAK44683 T-cell ep
28	32	88.9	33	7	ADL11905	ADL11905 Raveitoxi
29	32	88.9	43	4	AAW61570	AAW61570 Human imm
30	32	88.9	48	8	ABO57709	ABO57709 Human gen
31	32	88.9	53	4	AAW62764	AAW62764 Human imm
32	32	88.9	54	2	AAW05836	AAW05836 Mouse M-S
33	32	88.9	54	2	AAW59595	AAW59595 Mouse Ser
34	32	88.9	54	8	ADL17374	ADL17374 Mouse ser
35	32	88.9	71	5	ADK34982	ADK34982 Novel hum
36	32	88.9	74	4	AAU65187	AAU65187 Prolionib
37	32	88.9	74	4	ABG29902	ABG29902 Novel hum
38	32	88.9	74	6	ABM61706	ABM61706 Prolionib
39	32	88.9	88	6	ADA57035	ADA57035 Human sec
40	32	88.9	88	6	ADA40890	ADA40890 Human sec
41	32	88.9	88	6	ABR47812	ABR47812 Human sec
42	32	88.9	88	7	ADB91585	ADB91585 Human sec
43	32	88.9	88	7	ADD37916	ADD37916 Human sec
44	32	88.9	89	2	AAV38413	AAV38413 Human sec
45	32	88.9	89	2	AAV38428	AAV38428 Human sec
46	32	88.9	89	4	ABG03528	ABG03528 Novel hum
47	32	88.9	89	4	ABG03527	ABG03527 Novel hum
48	32	88.9	91	2	AAV35308	AAV35308 Chlamydia
49	32	88.9	91	4	AAU63229	AAU63229 Prolionib
50	32	88.9	91	6	ABM59748	ABM59748 Prolionib
51	32	88.9	108	4	AAU64030	AAU64030 Human sec
52	32	88.9	108	4	ABM63313	ABM63313 Drosophila
53	32	88.9	114	7	AD140621	AD140621 Human pur
54	32	88.9	116	4	ABG19371	ABG19371 Novel hum
55	32	88.9	122	4	AAU64490	AAU64490 Prolionib
56	32	88.9	122	6	ABM61009	ABM61009 Prolionib
57	32	88.9	126	6	AAU47152	AAU47152 Prolionib
58	32	88.9	126	6	ABM43671	ABM43671 Prolionib
59	32	88.9	135	7	ABO69120	ABO69120 Pseudomon
60	32	88.9	137	4	ABG19455	ABG19455 Novel hum
61	32	88.9	149	4	AAAB3271	AAAB3271 Human bre
62	32	88.9	149	4	AAAB3375	AAAB3375 Human bre
63	32	88.9	152	5	ABP58806	ABP58806 Calcium-d
64	32	88.9	171	7	ABO71893	ABO71893 Pseudomon
65	32	88.9	172	4	ABG11743	ABG11743 Novel hum
66	32	88.9	192	2	AAW18350	AAW18350 Prolifera
67	32	88.9	194	7	ADB64161	ADB64161 Human pro
68	32	88.9	194	7	ABO72453	ABO72453 Pseudomon
69	32	88.9	196	2	AAW73911	AAW73911 Mutant E2
70	32	88.9	197	4	ABG14123	ABG14123 Novel hum
71	32	88.9	208	4	AAW41818	AAW41818 Human pol
72	32	88.9	216	7	ADM05789	ADM05789 Human pro
73	32	88.9	226	7	ABO78982	ABO78982 Pseudomon
74	32	88.9	247	4	AAU31932	AAU31932 Novel hum
75	32	88.9	248	7	ADG75070	ADG75070 Human her
76	32	88.9	253	7	ADM04911	ADM04911 Human pro
77	32	88.9	264	7	ABO71628	ABO71628 Pseudomon
78	32	88.9	266	3	AAAG39368	AAAG39368 Arabidops
79	32	88.9	266	3	ADB64960	ADB64960 Human pro
80	32	88.9	272	3	AAAG42704	AAAG42704 Arabidops
81	32	88.9	275	2	AAV39302	AAV39302 Spm prot
82	32	88.9	275	2	AAV70948	AAV70948 S. spinos
83	32	88.9	275	6	ABP57683	ABP57683 Saccharop
84	32	88.9	276	3	AAAG39367	AAAG39367 Arabidops
85	32	88.9	280	6	ABR55197	ABR55197 Amino aci
86	32	88.9	283	4	ABG14124	ABG14124 Novel hum
87	32	88.9	284	4	AAW40032	AAW40032 Human pol
88	32	88.9	288	4	ABG26229	ABG26229 Novel hum
89	32	88.9	288	4	AAU29812	AAU29812 Novel hum
90	32	88.9	290	7	ABO74840	ABO74840 Pseudomon
91	32	88.9	296	6	ABP97797	ABP97797 Amino aci
92	32	88.9	302	6	ABR61804	ABR61804 Human JAG
93	32	88.9	302	5	ABR61734	ABR61734 Human JAG
94	32	88.9	343	5	ADK34132	ADK34132 Novel hum
95	32	88.9	351	7	ADB64772	ADB64772 Human pro
96	32	88.9	353	2	AAK75388	AAK75388 Japanese
97	32	88.9	353	2	AAK81587	AAK81587 Cedar pol
98	32	88.9	365	2	AAW73912	AAW73912 Mutant E2





PT obesity or diabetes.  
XX  
XX Claim 6; Page 232; 382pp; English.  
XX  
XX The invention relates to a complex between two interacting proteins in  
CC adipocyte cells, given in the specification. The proteins are identified  
CC by selecting a bait protein from a known adipocyte marker and then  
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by  
CC members of an adipocyte cDNA library. The proteins are designated SID  
CC (RTM) (selected interacting domains) proteins. Also included are a  
CC polynucleotide encoding a polypeptide in the adipocyte cells, a  
CC recombinant host cell expressing at least one of the interacting  
CC polypeptides of the complex, selecting a modulating compound in adipocyte  
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid  
CC sequences given in the specification (including its fragment or variant),  
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences  
CC given in the specification (including its fragment or variant), a vector  
CC comprising the SID (RTM) polynucleotide, a recombinant host cell  
CC comprising the vector, a protein chip comprising the polypeptides and a  
CC record comprising all or part of the data, listed in the specification.  
CC The complex, polypeptides, polynucleotides and compounds are useful for  
CC preventing or treating metabolic disorders such as obesity or diabetes.  
CC The polynucleotides are useful as probes or primers. The complex is  
CC particularly useful for identifying selected interacting domains (SID  
CC (RTM)) for screening drugs that modulate the protein interaction, thus  
CC exhibiting the therapeutic effect. The present sequence represents a SID  
CC (prey) protein of the invention  
XX  
XX Sequence 166 AA;  
SQ  
Query Match 100.0%; Score 36; DB 6; Length 166;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0;  
QY 1 AGGPGCV 6  
Db 39 AGGPGCV 44

RESULT 3  
ADJ69011  
ID ADJ69011 standard; protein; 407 AA.  
XX  
XX ADJ69011;  
AC  
XX 06-MAY-2004 (first entry)  
DT  
XX  
XX Human heart mitochondrial protein as a therapeutic target SeqID817.  
DE  
XX  
XX Mitochondrial; human; screening assay; diabetes mellitus;  
KM Huntington's disease; osteoarthritis;  
KM Leber's hereditary optic neuropathy; LHON;  
KM mitochondrial encephalopathy lactic acidosis and stroke; MELAS;  
KM myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;  
KM neuroprotective; noctropic; antidiabetic; anticonvulsant; antiarthritic;  
KM osteopathic; ophthalmological; cytostatic.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO2003087768-A2.  
PN  
XX  
XX 23-OCT-2003.  
PD  
XX  
XX 04-APR-2003; 2003WO-US010870.  
PF  
XX  
XX 12-APR-2002; 2002US-0372843P.  
PR 17-JUN-2002; 2002US-0389987P.  
PR 20-SEP-2002; 2002US-0412418P.  
XX  
XX (MITO-) MITOKOR.  
PA (BUCK-) BUCK INST AGE RES.  
XX  
XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM,

PI Warnock DE;  
XX  
XX WPI; 2003-845369/78.  
DR  
XX  
XX Identifying a mitochondrial target for drug screening assays and for  
PT treating diseases associated with altered mitochondrial function,  
PT comprises detecting a modified polypeptide in a sample and correlating  
PT with the disease.  
XX  
XX Claim 1; SEQ ID NO 817; 180pp; English.  
PS  
XX  
XX This invention relates to novel mitochondrial targets that can be used  
CC for therapeutic intervention in treating a disease associated with  
CC altered mitochondrial function. Specifically, it refers to a method for  
CC identifying proteins of the human heart mitochondrial proteome that are  
CC useful for drug screening assays, as well as therapeutic targets. The  
CC present invention describes a method for identifying such proteins that  
CC can be used in the treatment of various diseases associated with altered  
CC mitochondrial function including diabetes mellitus, Huntington's disease,  
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial  
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy  
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these  
CC compositions have neuroprotective, noctropic, antidiabetic,  
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and  
CC cytostatic activities. This polypeptide sequence is a human heart  
CC mitochondrial protein of the invention.  
XX  
XX Sequence 407 AA;  
SQ  
Query Match 100.0%; Score 36; DB 7; Length 407;  
Best Local Similarity 100.0%; Pred. No. 7.7e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0;  
QY 1 AGGPGCV 6  
Db 280 AGGPGCV 285

RESULT 4  
AAU27706  
ID AAU27706 standard; protein; 415 AA.  
XX  
XX AAU27706;  
AC  
XX 18-DEC-2001 (first entry)  
DT  
XX  
XX Human full-length polypeptide sequence #31.  
DE  
XX  
XX Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;  
KM mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukemia;  
KM cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;  
KM nervous system disorder; inflammatory disorder; cell differentiation;  
KM angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;  
KM genetic disorder; bone regeneration; tendon; ligament; tissue repair;  
KM cytostatic; antirheumatic; antiarthritic; vulnary; antiinflammatory;  
KM antibacterial; immunosuppressive; vasotropic; antiparkinsonian;  
KM neuroprotective; osteopathic; antidiabetic; antisthmatic; antiallergic;  
KM immunostimulant; analgesic; gene therapy.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200164834-A2.  
PN  
XX  
XX 07-SEP-2001.  
PD  
XX  
XX 26-FEB-2001; 2001WO-US004926.  
PF  
XX  
XX 28-FEB-2000; 2000US-00515126.  
PR 18-MAY-2000; 2000US-00577409.  
PR 17-JUN-2000; 2000US-00597707.  
PR 14-JUL-2000; 2000US-00616807.  
PR 19-SEP-2000; 2000US-00664641.  
XX

PA (HYSE-) HYSEQ INC.  
 XX Tang YF, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;  
 PI Drmanac R;  
 XX WPI; 2001-589862/66.  
 DR N-P8DB; AAS44606.  
 PT Novel polypeptides and nucleic acids obtained from cDNA libraries  
 prepared from various human tissues, for diagnosis, treatment of cancer,  
 PT neurological, inflammatory disorders and for use in arrays for detection.  
 XX  
 PS Claim 10; SEQ ID NO 203; 153bp; English.  
 CC Sequences AAU7676-AU28019 represent full-length polypeptides and contig  
 CC polypeptides of the invention. The proteins and their associated DNA  
 CC sequences are useful for the treatment, diagnosis and prevention of  
 CC various types of disorder in a mammalian subject such as a human, dog,  
 CC monkey, mouse, hamster or rat. The disorders include cancers such as  
 CC leukemia, lymphoma and neuroblastoma, autoimmune disorders such as  
 CC multiple sclerosis, connective tissue disease, rheumatoid arthritis,  
 CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system  
 CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's  
 CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and  
 CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's  
 CC disease, ischemia-reperfusion injury, shock, sepsis and inflammatory  
 CC bowel disease. The sequences exhibit activity relating to angiogenesis,  
 CC cell proliferation, cell differentiation, stem cell growth factor,  
 CC activin or inhibin. Therefore, they can be used to manipulate stem cells  
 CC in culture to give rise to neuroepithelial cells that can be used to  
 CC augment or replace cells damaged by illness, accidental damage or genetic  
 CC disorders. The sequences may also be used for regeneration of bone,  
 CC cartilage, tendons and ligaments and in tissue repair and burn healing.  
 CC Note: Some sequences for this patent did not form part of the printed  
 CC specification, but were obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 415 AA;  
 Query Match 100.0%; Score 36; DB 4; Length 415;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPCV 6  
 DB 288 AGGPCV 293  
 RESULT 5  
 ADJ58205  
 ID ADJ58205 standard; protein; 415 AA.  
 XX  
 AC ADJ58205;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Human Cezanne-2 C-terminal domain 455-858 SEQ ID NO:5.  
 XX  
 KW enzymatic deubiquitination; TRAFB domain;  
 KW tumour necrosis factor receptor associated factor binding domain;  
 KW Cezanne; core catalytic domain; human; Cezanne-2.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004011636-A2.  
 XX  
 PD 05-FEB-2004.  
 XX  
 PF 31-JUL-2003; 2003WO-GB003320.  
 XX  
 PR 31-JUL-2002; 2002US-0399439P.  
 XX  
 PR 25-MAR-2003; 2003GB-0006643.

XX  
 PA (BABR-) BABRAM INST.  
 XX  
 XX Evans PC, Kilshaw PJ;  
 PI  
 XX WPI; 2004-143859/14.  
 DR  
 XX  
 PT Enzymatic deubiquitination of a molecule which has been ubiquitinated to  
 PT form a ubiquitinated molecule, for stabilising a target molecule,  
 PT comprises contacting the ubiquitinated molecule with an enzyme homologous  
 PT to Cezanne.  
 XX  
 PS Claim 22; SEQ ID NO 5; 80bp; English.  
 CC  
 CC The present invention describes a method for enzymatic deubiquitination  
 CC of a molecule which has been ubiquitinated to form a ubiquitinated  
 CC molecule. The method comprises contacting the molecule with an enzyme or  
 CC domain with 20 % amino acid sequence identity with the TRAFB (tumour  
 CC necrosis factor (TNF) receptor associated factor binding) domain defined  
 CC by Cezanne amino acids 160-416 (ADJ58201), and/or the Cezanne core catalytic  
 CC Cezanne amino acids 126-455 (ADJ58204), and/or the Cezanne core catalytic  
 CC domain defined by Cezanne amino acids 182-455 (ADJ58206). Also described:  
 CC (1) a method for regulating the method described above by contacting a  
 CC molecule which has been ubiquitinated or a deubiquitination enzyme with a  
 CC regulatory polypeptide having at least 20% amino acid sequence identity  
 CC with the C-terminal domain of Cezanne defined by Cezanne amino acid  
 CC residues 444-858 (ADJ58205) or its functional fragment, homologue,  
 CC variant or mutant (for example, a conservatively substituted mutant); (2)  
 CC a method for modulating the activity of a molecule which is capable of  
 CC being ubiquitinated to form a ubiquitinated molecule, comprising  
 CC contacting the molecule or the ubiquitinated molecule with a polypeptide  
 CC cited above, where the polypeptide has deubiquitination activity; (3) a  
 CC protein having an amino acid sequence of Cezanne-2, or its functional  
 CC fragment, homologue, variant or mutant (for example, a conservatively  
 CC substituted mutant); and (4) an isolated nucleotide encoding the protein.  
 CC The methods are useful for stabilising the molecule or target molecule,  
 CC for targeting or maintaining the molecule or target molecule to or at a  
 CC sub-cellular location, and for modulating the activity of the molecule or  
 CC target molecule. The present sequence represents the human Cezanne-2 C-  
 CC terminal domain, which is used in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 415 AA;  
 Query Match 100.0%; Score 36; DB 8; Length 415;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPCV 6  
 DB 288 AGGPCV 293  
 RESULT 6  
 ABB97204  
 ID ABB97204 standard; protein; 479 AA.  
 XX  
 AC ABB97204;  
 XX  
 DT 28-JUN-2002 (first entry)  
 XX  
 DE Novel human protein SEQ ID NO: 472.  
 XX  
 KW Human; anti-inflammatory; vulnery; anti-inflammatory; immunomodulator;  
 KW anti-infectivity; cerebroprotective; cytoprotective; rheumatic; gene therapy;  
 KW neuroprotective; antiparkinsonian; protein therapy; BBT;  
 KW expressed sequence tag.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200222660-A2.  
 XX  
 PD 21-MAR-2002.

```

XX 10-SEP-2001; 2001WO-US026015.
XX
XX 11-SEP-2000; 2000US-00659671.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-292408/33.
XX N-PSDB; ABN32390.
XX
XX An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis.
XX
XX Example 2; SEQ ID NO 472; 509bp; English.
XX
XX The present invention provides the protein and coding sequences of 444
XX novel human proteins. These were isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX hematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
XX and cancer, to screen for drugs, to treat inflammatory conditions e.g.
XX rheumatoid arthritis, and to treat nervous system disorders e.g.
XX Parkinson's disease. The present sequence is a protein of the invention
XX
XX Sequence 479 AA;
SQ
Query Match          100.0%; Score 36; DB 5; Length 479;
Best Local Similarity 100.0%; Pred. NO. 8.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 AGGRCV 6
        |||||
Db       352 AGRCV 357

RESULT 7
AD126089
ID AD126089 standard; protein; 843 AA.
XX
XX AD126089;
XX
XX 22-APR-2004 (first entry)
XX
XX Human protein that promotes STAT6 activation #27.
DE
XX human; signal transducer and activator of transcription 6; STAT6;
XX immunogen; STAT6 activation; allergy; inflammation; autoimmune disease;
XX diabetes; hyperlipidaemia; infection; cancer; Th1 hyperactive disease;
XX rheumatoid arthritis; osteoarthritis; systemic lupus erythematosus;
XX sepsis; asthma; allergic rhinitis; ischemic heart disease;
XX eubacterial haemorrhage; viral hepatitis; AIDS.
XX
XX Homo sapiens.
XX
XX WO2003104277-A2.
XX
XX 18-DEC-2003.
XX
XX 05-JUN-2003; 2003WO-JP007123.
XX
XX 05-JUN-2002; 2002JP-00164257.
XX 06-JUN-2002; 2002US-0385912P.
XX 26-DEC-2002; 2002JP-00377326.
XX 27-DEC-2002; 2002US-0436467P.
XX 15-MAY-2003; 2003JP-00137505.
XX 16-MAY-2003; 2003US-0470836P.
XX
XX (ASAH) ASAH KASEI KK.

```

```

XX Sugahara T, Matsuda A, Honda G, Muramatsu S, Ishizawa K;
XX WPI; 2004-122214/12.
XX N-PSDB; AD126088.
XX
XX New signal transducer and activator of transcription 6 activation
XX promoting purified protein, for diagnosing and treating disease
XX associated with activation/inhibition of transcription factor e.g.
XX diabetes and cancer.
XX
XX Claim 1; SEQ ID NO 54; 1368bp; English.
XX
XX The invention relates to a purified protein promoting signal transducer
XX and activator of transcription 6 activation (STAT6). The protein is
XX useful for the producing an antibody, which involves administering the
XX protein or its epitope-bearing fragments to a non-human animal as an
XX antigen. The nucleic acid is useful for diagnosing a disease or
XX susceptibility to a disease related to expression or activity of the
XX protein. A transformant expressing the protein is useful for screening
XX compounds which inhibit or promote STAT6 activation. A transformant
XX expressing the protein is useful for producing a pharmaceutical
XX composition. Compositions, antibodies and antisense molecules are useful
XX for the treating a disease associated with STAT6 activation such as
XX allergic diseases, inflammation, autoimmune diseases, diabetes,
XX hyperlipidaemia, infectious disease and cancers. Compositions are useful
XX for treating diseases associated with STAT6 activation and/or prevention
XX of Th1 hyperactive diseases. Compositions are also useful in rheumatoid
XX arthritis, osteoarthritis, systemic lupus erythematosus, sepsis, asthma,
XX allergic rhinitis, ischemic heart disease, eubacterial haemorrhage,
XX viral hepatitis and AIDS. The protein has efficient promoting STAT6
XX activity. The protein or nucleic acid is effectively useful for screening
XX compounds for treating and preventing disease associated with excessive
XX activation or inhibition of STAT6. The present sequence represents the
XX amino acid sequence of a human protein which promotes STAT6 activation.
XX
XX Sequence 843 AA;
SQ
Query Match          100.0%; Score 36; DB 8; Length 843;
Best Local Similarity 100.0%; Pred. NO. 1.4e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 AGGRCV 6
        |||||
Db       716 AGRCV 721

RESULT 8
AD126087
ID AD126087 standard; protein; 843 AA.
XX
XX AD126087;
XX
XX 22-APR-2004 (first entry)
XX
XX Human protein that promotes STAT6 activation #26.
DE
XX human; signal transducer and activator of transcription 6; STAT6;
XX immunogen; STAT6 activation; allergy; inflammation; autoimmune disease;
XX diabetes; hyperlipidaemia; infection; cancer; Th1 hyperactive disease;
XX rheumatoid arthritis; osteoarthritis; systemic lupus erythematosus;
XX sepsis; asthma; allergic rhinitis; ischemic heart disease;
XX eubacterial haemorrhage; viral hepatitis; AIDS.
XX
XX Homo sapiens.
XX
XX WO2003104277-A2.
XX
XX 18-DEC-2003.
XX
XX 05-JUN-2003; 2003WO-JP007123.
XX
XX 05-JUN-2002; 2002JP-00164257.
XX

```

PR 06-JUN-2002; 2002US-0385912P.  
 PR 26-DEC-2002; 2002JP-00377326.  
 PR 27-DEC-2002; 2002JP-0036467P.  
 PR 15-MAY-2003; 2003JP-00137505.  
 PR 16-MAY-2003; 2003US-0470836P.  
 PA (ASAH) ASAH KASEI KK.  
 PI Sugahara T, Matsuda A, Honda G, Muramatsu S, Ishizawa K;  
 DR WPI; 2004-122214/12.  
 DR N-PSDB; ADI26086.  
 XX  
 PT New signal transducer and activator of transcription 6 activation  
 PT promoting purified protein, for diagnosing and treating disease  
 PT associated with activation/inhibition of transcription factor e.g.  
 PT diabetes and cancer.  
 PS Claim 1, SEQ ID NO 52; 1368pp; English.  
 XX  
 CC The invention relates to a purified protein promoting signal transducer  
 CC and activator of transcription 6 activation (STAT6). The protein is  
 CC useful for the producing an antibody, which involves administering the  
 CC protein or its epitope-bearing fragments to a non-human animal as an  
 CC antigen. The nucleic acid is useful for diagnosing a disease or  
 CC susceptibility to a disease related to expression or activity of the  
 CC protein. A transformant expressing the protein is useful for screening  
 CC compounds which inhibit or promote STAT6 activation. A pharmaceutical  
 CC composition. Compositions, antibodies and antisense molecules are useful  
 CC for the treating a disease associated with STAT6 activation such as  
 CC allergic diseases, inflammation, autoimmune diseases, diabetes,  
 CC hyperlipidemia, infectious diseases and cancers. Compositions are useful  
 CC for treating disease associated with STAT6 activation and/or prevention  
 CC of Th1 hyperactive diseases. Compositions are also useful in rheumatoid  
 CC arthritis, osteoarthritis, systemic lupus erythematosus, sepsis, asthma,  
 CC allergic rhinitis, ischemic heart diseases, subarachnoid hemorrhage,  
 CC viral hepatitis and AIDS. The protein has efficient promoting STAT6  
 CC activity. The protein or nucleic acid is effectively useful for screening  
 CC compounds for treating and preventing disease associated with excessive  
 CC activation or inhibition of STAT6. The present sequence represents the  
 CC amino acid sequence of a human protein which promotes STAT6 activation.  
 XX  
 SQ Sequence 843 AA;  
 Query Match 100.0%; Score 36; DB 8; Length 843;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPCV 6  
 Db 716 AGGPCV 721  
 RESULT 9  
 ABR47408  
 ID ABR47408 standard; protein; 858 AA.  
 AC ABR47408;  
 DT 12-JUN-2003 (first entry)  
 XX  
 DE Breast cancer associated protein sequence SEQ ID NO:47.  
 XX  
 KW Human; breast cancer; cytostatic; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003004989-A2.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 21-JUN-2002; 2002WO-US019669.

XX 21-JUN-2001; 2001US-0299887P.  
 PR 27-JUN-2001; 2001US-0301572P.  
 PR 18-JUL-2001; 2001US-0306501P.  
 PR 25-SEP-2001; 2001US-0325002P.  
 PR 05-MAR-2002; 2002US-0362585P.  
 PR 14-MAY-2002; 2002US-0380391P.  
 XX  
 PA (MILL-) MILLENIUM PHARM INC.  
 XX  
 PI Little J, Gannavarapu M, Glatt K, Hoersch S, Kamatkar S;  
 PI Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RS,  
 PI Bast RC, Horobagyl GN, Pusztai L, Meric F, Sahin A, Mills GB;  
 DR WPI; 2003-210381/20.  
 DR N-PSDB; ACC50099.  
 XX  
 PT Breast cancer diagnosis or treatment by comparing the level of expression  
 PT of a marker in a patient sample with that in the control non-breast  
 PT cancer sample.  
 PS Claim 1, SEQ ID NO 47; 128pp; English.  
 XX  
 CC The present invention describes a method for assessing whether a patient  
 CC is afflicted with breast cancer. The method comprises comparing the level  
 CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and  
 CC ABR4736 to ABR47632) in a patient sample and the normal level of  
 CC expression of the marker in a control non-breast cancer sample, where a  
 CC significant increase in the level of expression of the marker in the  
 CC patient sample and the normal level is an indication that the patient is  
 CC afflicted with breast cancer. The breast cancer associated sequences from  
 CC the present invention have cytostatic activities and can be used in gene  
 CC therapy. The method is useful for diagnosing and treating breast cancer.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 858 AA;  
 Query Match 100.0%; Score 36; DB 6; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPCV 6  
 Db 731 AGGPCV 736  
 RESULT 10  
 ADC37584  
 ID ADC37584 standard; protein; 858 AA.  
 AC ADC37584;  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE Human nucleic acid associated protein, NAAp-51.  
 XX  
 KW Human; nucleic acid associated protein; NAAp; cytostatic;  
 KW antiarteriosclerotic; anticonvulsant; neurotropic; neuroprotective;  
 KW cerebroprotective; anti-HIV; anti-allergic; anti-inflammatory;  
 KW thymometric; gene therapy; cell proliferative disorder; cancer;  
 KW atherosclerosis; neurological disorder; epilepsy; Huntington's disease;  
 KW stroke; immune disorder; inflammatory disorder; AIDS; allergy;  
 KW developmental disorder; Hypothyroidism; Cushing's syndrome; infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003046151-A2.  
 XX  
 PD 05-JUN-2003.  
 XX  
 PF 26-NOV-2002; 2002WO-US038445.

XX	27-NOV-2001; 2001US-0333925P.
PR	07-DEC-2001; 2001US-0340477P.
PR	14-DEC-2001; 2001US-0340362P.
PR	18-DEC-2001; 2001US-0342002P.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Baughn MR, Becha SD, Bhactia U, Blake JU, Burford N, Burrill JD,
PI	Chang H, Chawla NK, Elliott VS, Emerling BM, Forsythe TJ, Gandhi AR,
PI	Gleizen KJ, Goryard AE, Griffin JA, Hatfield AJA, Jackson JL, Ho A,
PI	Ison CH, Jackson AA, Jiang X, Jin P, Kabie AE, Khare R, Lal PG,
PI	Lee EA, Lee S, Lee SY, Li JK, Lu DM, Ramkumar J, Richardson TW,
PI	Sprague KW, Swarnakar A, Tang YT, Warren BA, Xu Y, Yao MG, Yue H,
PI	Zheng W,
DR	WPI; 2003-513642/48.
DR	N-PSDB; ADC37644.
XX	
PT	New human nucleic acid associated proteins (NAAP), useful for diagnosing,
PT	treating and preventing diseases or conditions associated with the
PT	aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
PT	infections.
XX	
PS	Claim 1; SEQ ID NO 51; 383bp; English.
XX	
CC	The present invention relates to novel human nucleic acid associated
CC	protein (NAAP) (1; ADC37534-ADC37593) and their coding sequences ( ). The
CC	NAAPs and their coding sequences are useful in diagnosing, treating and
CC	preventing diseases or conditions associated with the decreased
CC	expression or over expression of NAAP, such as cell proliferative (e.g.
CC	cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's
CC	disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and
CC	developmental (e.g. Hypothyroidism, Cushing's syndrome) disorders, or
CC	infections. These are also useful in assessing the effects of exogenous
CC	compounds on the expression of nucleic acid and amino acid sequences of
CC	NAAP.
XX	
XX	Sequence 858 AA;
XX	
Qy	1 AGGPGCV 6
Db	731 AGGPGCV 736
RESULT 11	
ID	ADP76869
XX	ADF76869 standard; protein; 1745 AA.
XX	
AC	ADF76869;
XX	
DT	26-FEB-2004 (first entry)
XX	
DE	Novel human secreted and transmembrane protein Seqid 544.
XX	
KW	human, PRO; membrane bound protein; membrane bound receptor;
KW	cell proliferation; cell migration; cell differentiation;
KW	mitogenic factor; survival factor; cytotoxic factor;
KW	differentiation factor; neuropeptide; hormone; cell receptor;
KW	receptor-ligand interaction; cytosstatic; chondrocyte; tumour.
XX	
OS	Homo sapiens.
XX	
XX	WO2003072035-A2.
XX	
PD	04-SEP-2003.
XX	
XX	21-FEB-2003; 2003WO-US005241.
XX	

```

PR 22-FEB-2002; 2002US-0359461P.
PA
XX (GETH ) GENENTECH INC.
XX
XX Bodary SC, Clark H, Hunte B, Jackman JK, Schoenfeld JR;
PI Williams PM, Wood WI, Wu TD;
XX
XX WPI, 2003-721702/68.
DR N-PDB; ADF76868.
XX
XX New PRO polypeptides, useful for diagnosing and treating an immune
PT related disorder, e.g. systemic lupus erythematosus, rheumatoid
PT arthritis, osteoarthritis, juvenile chronic arthritis, thyroiditis or
PT diabetes mellitus.
XX
XX Claim 10; SEQ ID NO 544; 918pp; English.
XX
XX This invention relates to novel nucleic acids encoding human PRO secreted
CC and transmembrane proteins. Extracellular proteins play important roles
CC in the formation, differentiation and maintenance of multicellular
CC organisms. The fate of many individual cells (for example proliferation,
CC migration or differentiation) is typically governed by information
CC received from other cells and the immediate environment. The information
CC is often transmitted by secreted polypeptides (for example mitogenic
CC factors, survival factors, cytotoxic factors, differentiation factors,
CC neuropeptides and hormones) which are received and interpreted by diverse
CC cell receptors or membrane bound proteins. These membrane bound proteins
CC and receptors may be of use as pharmaceutical and diagnostic agents, such
CC as in the blocking of receptor-ligand interactions. The current invention
CC provides the amino acid sequences of novel human membrane bound receptors
CC and proteins, along with the cDNA sequences encoding them. The novel
CC proteins of the invention may have cytosolic activities through the
CC stimulation of chondrocytes. The nucleic acids of the invention may be
CC useful for the manufacture of a medicament for diagnosing or treating a
CC tumour in a mammal. In addition, they may be useful for measuring or
CC detecting the expression of a tumour associated gene. The present
CC sequence is the amino acid sequence of a human PRO protein of the
CC invention.
XX
XX Sequence 1745 AA:
SQ
XX
XX Query Match 100.0%; Score 36; DB 7; Length 1745;
XX Best Local Similarity 100.0%; Pred. No. 2.7e+03;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
XX
XX QY 1 AGGRCV 6
XX |||||
XX 1 AGGRCV 6
XX
XX Db
XX
XX RESULT 12
XX ADO71781
XX ID ADO71781 standard; protein; 1745 AA.
XX
XX AC ADO71781;
XX
XX DT 26-AUG-2004 (first entry)
XX
XX DE BFLP0169 related protein, SEQ ID 21.
XX
XX KW Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;
XX kidney; lupus nephritis; autoimmune disorder; lupus.
XX
XX OS Homo sapiens.
XX
XX FN WO2004048521-A2.
XX
XX PD 10-JUN-2004.
XX
XX PP 21-NOV-2003; 2003WO-US037339.
XX
XX PR 21-NOV-2002; 2002US-0428094P.
XX

```

PA (AMHP) WYETH.  
 XX O'loole M, Mounts WM, Shojae N;  
 XX  
 DR WPI; 2004-441163/41.  
 XX  
 PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 XX or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 XX  
 PS Disclosure; Page 9-13; 83pp; English.  
 XX  
 CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC was used in a sequence alignment with the BFLP0169 protein.  
 XX  
 SQ Sequence 1745 AA;  
 Query Match 100.0%; Score 36; DB 8; Length 1745;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPGCV 6  
 DB 1 AGGPGCV 6  
 AC ADO71782;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE BFLP0169 related protein, SEQ ID 22.  
 XX  
 KW Antiinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 XX kidney; lupus nephritis; autoimmune disorder; lupus.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004046521-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 21-NOV-2003; 2003WO-US037339.  
 XX  
 PR 21-NOV-2002; 2002US-0428094P.  
 XX  
 PA (AMHP) WYETH.  
 XX  
 PI O'loole M, Mounts WM, Shojae N;  
 XX  
 DR WPI; 2004-441163/41.  
 XX  
 PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 XX or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 XX  
 PS Disclosure; Page 9-13; 83pp; English.  
 XX  
 CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC was used in a sequence alignment with the BFLP0169 protein.  
 XX  
 SQ Sequence 1767 AA;

Query Match 100.0%; Score 36; DB 8; Length 1767;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPGCV 6  
 DB 19 AGGPGCV 24  
 AC ABOG09906;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #9897.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 XX  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSB-) HYSBQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 XX  
 DR N-PSDB; AAS74093.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 XX responsible for genetic disorders or other traits and to assess  
 XX biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 40265; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABOG0010-ABG0377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1889 AA;  
 Query Match 100.0%; Score 36; DB 4; Length 1889;

Best Local Similarity 100.0%; Pred. No. 2.9e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6  
DB 1 AGGPCV 6

## RESULT 15

AGG79569  
ID AGG79569 standard; peptide, 142 AA.

XX AAG79569;

DT 23-DEC-2002 (first entry)

XX ACK1 C-terminal serine/proline-rich region.

XX Mlg-6; lung; fibroblast; gene 33; rat; hyperproliferation;

KM serine/threonine phosphorylation site; inflammation; tumor; Erk2;

KM Cdc42 and rac interaction binding; CRIB; SH3; WW; binding motif;

KM tyrosine kinase; ACK1; epidermal growth factor receptor; EGFR;

XX Rattus rattus.

PN EPI236474-A1.

PD 04-SEP-2002.

PF 26-FEB-2001; 2001EP-00104737.

PR 26-FEB-2001; 2001EP-00104737.

PI (PIAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

DR WPI, 2002-692997/75.

PT Modulation of epidermal growth factor receptor activity and treatment of  
hyperproliferative diseases, inflammation and tumors comprises  
administration of a Mlg-6 protein.

PS Example; Fig 1B; 25pp; English.

XX This sequence represents the C-terminal serine- and proline-rich region  
of ACK1. This sequence has 71% sequence homology with the C-terminal  
region of Mlg-6. A Mlg-6 was first cloned in 1995 by screening a library  
of lung fibroblasts which had been stimulated with fetal calf serum for  
inducing genes. Mlg-6 shows 84% homology with rat gene 33 product and has  
several potential serine/threonine phosphorylation sites, a potential  
Cdc42 and rac interaction binding (CRIB) domain, several SH3 and WW  
domain binding motifs and this serine and proline-rich carboxy terminal  
domain. Mlg-6 can be used for manufacture of an agent for modulation of  
epidermal growth factor receptor (EGFR) activity. The Mlg-6 protein is  
useful for modulation of EGFR activity, especially inhibition of the  
mitogen-activated protein kinase pathway (inhibition of Erk2 activity),  
and for treatment of EGFR overexpression associated disorders (especially  
hyperproliferative diseases, inflammation and tumors). Mlg-6 protein and  
nucleic acids encoding for it are also useful as a target for modulation  
of EGFR activity

XX Sequence 142 AA;

QY Query Match 97.2%; Score 35; DB 5; Length 142;

Best Local Similarity 83.3%; Pred. No. 4.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6  
DB 109 AGGPCT 114

RESULT 16  
ABR82256  
ID ABR82256 standard; protein, 1036 AA.

XX ABR82256;

DT 13-OCT-2003 (first entry)

XX Human activated p12cdc42hs kinase (ACK1).

KM ACK1; activated p12cdc42hs kinase; cytosolic; gene therapy; human;

KM immune response inducer; cancer; enzyme.

XX Homo sapiens.

PN WO2003054512-A2.

PD 03-JUL-2003.

PF 13-DEC-2002; 2002MO-US039927.

PR 20-DEC-2001; 2001US-0341436P.

PA (TULIA-) TULARIK INC.

PI Degenhardt YV, Powers S;

DR WPI, 2003-569275/53.

DR N-PSDB; ACC85084.

PT Diagnosing cancer e.g., ovarian cancer, prostate cancer, breast cancer in  
a mammal or inhibiting cancer or precancerous growth in a mammalian  
tissue, using activated p12cdc42hs kinase gene or gene product.

PS Claim 12; Page 95; 100pp; English.

XX The invention relates to diagnosing cancer in a mammal or inhibiting  
cancer or precancerous growth in a mammalian tissue, using activated  
p12cdc42hs kinase (ACK1) gene or gene product. The methods are useful for  
diagnosing cancer e.g., breast cancer, prostate cancer or ovarian cancer  
in a mammal. Methods are provided for inhibiting cancer or precancerous  
growth in a mammalian tissue such as breast tissue, prostate tissue or  
ovarian tissue which involves contacting the tissue with an inhibitor  
that interacts with ACK1 DNA or RNA and thus inhibits ACK1 gene function,  
or contacting the tissue with an inhibitor of ACK1 protein. Administering  
siRNA to a patient, where the siRNA interacts with ACK1 gene or ACK1 mRNA  
transcript, is useful for blocking in vivo expression of a gene. The  
siRNA interferes with ACK1 activity or causes post-transcriptional  
silencing of ACK1 gene in a mammalian cell e.g., human cell. An isolated  
ACK1 gene amplicon is useful in gene therapy, and for inducing an immune  
response in a mammal. The present sequence represents a human ACK1  
sequence (GenBank Accession no. NP\_005772.2)

XX Sequence 1036 AA;

QY Query Match 97.2%; Score 35; DB 6; Length 1036;

Best Local Similarity 83.3%; Pred. No. 2.5e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6  
DB 837 AGGPCT 842

## RESULT 17

AD808157  
ID AD808157 standard; protein, 1068 AA.

XX AD808157;

DT 29-JAN-2004 (first entry)







PT New peptide compounds useful for detecting herpes simplex virus type 2 -  
 PT can differentiate between herpes simplex virus type 1 and type 2.

PS Example 1, Page 10; 25pp; English.

XX The invention relates to a multiply displayed peptide structure of  
 CC formula [(X1)p-A-(X2)q-SP]-n-Core, X1, X2 = 1-6 non-interfering amino acid  
 CC residues; A = Glu Gly Phe Gly Ala Gly Asp Gly Pro Gly Asp Asp  
 CC Asp; Sp = spacer group extending outwardly from the core; n at least 4;  
 CC and p, q = 0 or 1. The linkage between the core and the spacer may be  
 CC chemical or physical. Peptides of this formula are used in the diagnosis  
 CC of herpes simplex virus type 2 (HSV-2). The peptides can distinguish HSV-  
 CC 2 from HSV-1. A series of 67 peptides (AAW66624-W66690), mostly 18 amino  
 CC acids long, that spanned amino acids 21-699 of the predicted open reading  
 CC frame of HSV gG2 were synthesized. The peptides were made as multiply  
 CC displayed peptide structures of the invention and were screened against  
 CC sera from HSV-1, HSV-2 antibody-positive individuals and from individuals  
 CC having no laboratory evidence of HSV infection. From the results, peptide  
 CC 55 (AAW66678) was considered a likely candidate for type specific  
 CC serodiagnosis of HSV

XX Sequence 18 AA;

Query Match 88.9%; Score 32; DB 2; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
 |||||  
 DB 5 GGPCV 9

RESULT 20

AAK45552  
 ID AAK45552 standard; protein, 20 AA.

XX AAK45552;

AC 25-MAR-2003 (revised)  
 DT 13-JUL-1994 (first entry)

XX Cry j I pollen allergen peptide CJI-11.

XX Japanese cedar; detection; allergy; treatment; diagnosis; T cell epitope;  
 KM sensitivity.

XX Cryptomeria japonica.

XX WO9401560-A1.

XX 20-JAN-1994.

XX 15-JAN-1993; 93WO-US000139.

XX 10-JUL-1992; 92WO-US005661.

XX 01-SEP-1992; 92US-00938990.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

XX Griffith IU, Pollock J, Bond JF, Garman RD, Kuo M;

XX WPI, 1994-035066/04.

XX Antigens derived from Japanese cedar pollen allergen Cry j I - contain at  
 PT least two T cell epitope(s), used to treat or diagnose allergy.

XX Claim 1; Fig 13; 137pp; English.

XX The sequence is that of an isolated peptide of the Japanese cedar pollen  
 CC allergen Cry j I (amino acids 101-120). The peptide, CJI-11, can be used  
 CC for the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced side  
 CC effects compared to naturally occurring allergens. (Updated on 25-MAR-

CC 2003 to correct PN field.)

XX Sequence 20 AA;

Query Match 88.9%; Score 32; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
 |||||  
 DB 4 GGPCV 8

RESULT 21

AAK45551  
 ID AAK45551 standard; protein, 20 AA.

XX AAK45551;

AC 25-MAR-2003 (revised)  
 DT 13-JUL-1994 (first entry)

XX Cry j I pollen allergen peptide CJI-10.

XX Japanese cedar; detection; allergy; treatment; diagnosis; T cell epitope;  
 KM sensitivity.

XX Cryptomeria japonica.

XX WO9401560-A1.

XX 20-JAN-1994.

XX 15-JAN-1993; 93WO-US000139.

XX 10-JUL-1992; 92WO-US005661.

XX 01-SEP-1992; 92US-00938990.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

XX Griffith IU, Pollock J, Bond JF, Garman RD, Kuo M;

XX WPI, 1994-035066/04.

XX Antigens derived from Japanese cedar pollen allergen Cry j I - contain at  
 PT least two T cell epitope(s), used to treat or diagnose allergy.

XX Claim 1; Fig 13; 137pp; English.

XX The sequence is that of an isolated peptide of the Japanese cedar pollen  
 CC allergen Cry j I (amino acids 91-110). The peptide, CJI-10, can be used  
 CC for the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced side  
 CC effects compared to naturally occurring allergens. (Updated on 25-MAR-  
 CC 2003 to correct PN field.)

XX Sequence 20 AA;

Query Match 88.9%; Score 32; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
 |||||  
 DB 14 GGPCV 18

RESULT 22

AAK82501  
 ID AAK82501 standard; protein, 20 AA.

XX AAK82501;

XX



CC (Cryptomeria japonica) pollen allergen Cry j 1 which stimulate an immune  
 CC response in dogs. The invention also encompasses DNA encoding the Cry j 1  
 CC T-cell epitopes, a vector comprising a Cry j 1 T-cell epitope-encoding  
 CC DNA, and DNA vaccine containing a vector of the invention for the  
 CC treatment of allergic dermatitis. The Cry j 1 T-cell epitopes and their  
 CC encoding nucleic acids are useful in the preparation of vaccines,  
 CC especially DNA vaccines, for the treatment of allergic dermatitis in  
 CC dogs. The DNA vaccine of the invention can be prepared inexpensively with  
 CC easier manufacturing and purification processes. The present sequence  
 CC represents a specifically claimed Cry j 1 T-cell epitope of the  
 CC invention.

XX  
 XX Sequence 20 AA;

Query Match 88.9%; Score 32; DB 7; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
 |||||  
 Db 4 GGPCV 8

RESULT 25

ADD93844 standard; peptide; 20 AA.

ADD93844;

29-JAN-2004 (first entry)

Japanese cedar pollen allergen Cry j 1 T-cell epitope. SEQ ID NO:10.

KW Japanese cedar; pollen allergen; Cry j 1; T-cell epitope; dog;  
 KM Canis familiaris; allergic dermatitis; veterinary; DNA vaccine;  
 XX dermatological; antiinflammatory; vaccine; gene therapy.

OS Cryptomeria japonica.

PN JP2003116556-A.

PD 22-APR-2003.

PF 09-OCT-2001; 2001JP-00311433.

PR 09-OCT-2001; 2001JP-00311433.

PA (NIPZ) NIPPON ZENYAKU KOGYO KK.

DR WPI; 2003-771273/73.

PT New cedar pollinosis antigen T-cell epitope in a dog useful for preparing  
 PT a DNA vaccine that can be used to prevent allergic dermatitis.

PS Claim 1; SEQ ID NO 10; 20pp; Japanese.

XX The invention relates to T-cell epitopes of the Japanese cedar  
 CC (Cryptomeria japonica) pollen allergen Cry j 1 which stimulate an immune  
 CC response in dogs. The invention also encompasses DNA encoding the Cry j 1  
 CC T-cell epitopes, a vector comprising a Cry j 1 T-cell epitope-encoding  
 CC DNA, and DNA vaccine containing a vector of the invention for the  
 CC treatment of allergic dermatitis. The Cry j 1 T-cell epitopes and their  
 CC encoding nucleic acids are useful in the preparation of vaccines,  
 CC especially DNA vaccines, for the treatment of allergic dermatitis in  
 CC dogs. The DNA vaccine of the invention can be prepared inexpensively with  
 CC easier manufacturing and purification processes. The present sequence  
 CC represents a specifically claimed Cry j 1 T-cell epitope of the  
 CC invention.

XX Sequence 20 AA;

Query Match 88.9%; Score 32; DB 7; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 2 GGPCV 6  
 |||||  
 Db 14 GGPCV 18

RESULT 26

AAR45579 standard; protein; 30 AA.

AAR45579;

25-MAR-2003 (revised)

DT 13-JUL-1994 (first entry)

DE Cry j 1 pollen allergen peptide CJI-41.

KW Japanese cedar; detection; allergy; treatment; diagnosis; T cell epitope;  
 KM sensitivity.

OS Cryptomeria japonica.

PN WO9401560-A1.

PD 20-JAN-1994.

PF 15-JAN-1993; 93MO-US000139.

PR 10-JUL-1992; 92MO-US0005661.

PR 01-SEP-1992; 92US-00938990.

XX (TMMU-) IMMUNOLOGIC PHARM CORP.

XX Griffith J, Pollock J, Bond JF, Garman RD, Kuo M;  
 XX WPI; 1994-035066/04.

PT Antigens derived from Japanese cedar pollen allergen Cry j 1 - contain at  
 PT least two T cell epitope(s), used to treat or diagnose allergy.

PS Claim 76; Fig 18; 137pp; English.

CC The sequence is that of an isolated peptide of the Japanese cedar pollen  
 CC allergen Cry j 1. The peptide, CJI-41, can be used for the treatment and  
 CC diagnosis of allergies associated with Japanese cedar pollen. It has  
 CC enhanced therapeutic properties but reduced side effects compared to  
 CC naturally occurring allergens. (Updated on 25-MAR-2003 to correct PN  
 CC field.)

XX Sequence 30 AA;

Query Match 88.9%; Score 32; DB 2; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
 |||||  
 Db 24 GGPCV 28

RESULT 27

AAM44683 standard; peptide; 30 AA.

AAM44683;

01-MAY-1998 (first entry)

XX T-cell epitope peptide #2 of sugi pollen antigen.

XX T-cell epitope, sugi pollen antigen; sugi pollinosis.

OS Synthetic.  
OS Cryptomeria japonica.  
XX  
PN JP10007700-A.  
XX  
PD 13-JAN-1998.  
XX  
PF 24-JUN-1996; 96JP-00163287.  
XX  
PR 24-JUN-1996; 96JP-00163287.  
XX  
PA (MEIJ ) MEIJI SEIKA KAISHA LTD.  
PA (DAIJI ) DAICEL CHEM IND LTD.  
XX  
DR WPI; 1998-133630/13.  
XX  
PT T cell epitope peptide of sugi pollen antigen - useful in the treatment  
PT of sugi pollinosis.  
XX  
PS Claim 1; Page 4; 14pp; Japanese.  
XX  
CC T-cell epitope peptides AAM4682-88 and their derivatives react with sugi  
CC pollinosis patient peripheral blood T lymphocytes. A composition prepared  
CC by combining at least 2 of the above peptides and/or their derivatives is  
CC used for the prevention and treatment of sugi pollinosis  
XX  
SQ Sequence 30 AA;  
Query Match 88.9%; Score 32; DB 2; Length 30;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 2 GSPCV 6  
DB 4 GSPCV 8  
RESULT 28  
ADL11905  
ID ADL11905 standard; protein; 33 AA.  
XX  
AC ADL11905;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
DE Ravenoxin-II protein.  
XX  
KW Bacillus thuringiensis; spider toxin gene; biopesticide.  
XX  
OS Bacillus thuringiensis.  
XX  
PN CN1366822-A.  
XX  
PD 04-SEP-2002.  
XX  
PF 18-JUL-2001; 2001CN-00114592.  
XX  
PR 18-JUL-2001; 2001CN-00114592.  
XX  
PA (UYHU-) UNIV HUNAN NORMAL.  
XX  
PI Xia L, Liang S, Ding X;  
XX  
DR WPI; 2003-483110/46.  
DR N-Psdb; ADL11906.  
XX  
PT New strain of Bacillus thuringiensis, containing a spider toxin gene and  
PT a promoter sequence, is used as a biopesticide.  
XX  
PS Disclosure; SEQ ID NO 9; 28pp; Chinese.  
XX  
CC The present invention relates to a Bacillus thuringiensis strain  
CC comprising a spider toxin gene and a strong promoter sequence. The

CC B.thuringiensis is used as a biopesticide as it can produce the  
CC B.thuringiensis toxin and a spider toxin. The present sequence represents  
CC Ravenoxin-II protein.  
XX  
SQ Sequence 33 AA;  
Query Match 88.9%; Score 32; DB 7; Length 33;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 2 GSPCV 6  
DB 6 GSPCV 10  
RESULT 29  
AAM91570  
ID AAM91570 standard; protein; 43 AA.  
XX  
AC AAM91570;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human Immune/haematopoietic antigen SEQ ID NO:19163.  
XX  
KW Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis.  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001354.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214866P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227109P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0233397P.  
 PR 14-SEP-2000; 2000US-0233398P.  
 PR 14-SEP-2000; 2000US-0233399P.  
 PR 14-SEP-2000; 2000US-0234001P.  
 PR 14-SEP-2000; 2000US-023401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235835P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239335P.  
 PR 13-OCT-2000; 2000US-0239337P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241825P.  
 PR 01-NOV-2000; 2000US-024617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.

PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251858P.  
 PR 08-DEC-2000; 2000US-0251988P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR MPI; 2001-483426/52.  
 XX  
 DR N-PSDB; AAK64351.  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 useful for preventing, diagnosing and/or treating cancers and metastasis.  
 XX  
 PS Claim 11; SEQ ID NO 19163; 3071pp + Sequence listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK67694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169  
 CC represent sequences used in the exemplification of the present invention  
 XX  
 SQ Sequence 43 AA;  
 Query Match 88.9%; Score 32; DB 4; Length 43;  
 Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GPCV 6  
 |||||  
 Db 2 GPCV 6  
 RESULT 30  
 ABO57709  
 ID ABO57709 standard; protein; 48 AA.  
 XX  
 AC ABO57709;

XX 29-JUL-2004 (first entry)  
DT Human genome derived single exon protein #3943.  
XX  
DB Human; gene expression; single exon probe; microarray;  
XX alternative splicing event; genomic alteration.  
KW Homo sapiens.  
XX US2003194704-A1.  
XX  
XX 16-OCT-2003.  
PD  
XX 03-APR-2002; 2002US-00029386.  
XX  
XX 03-APR-2002; 2002US-00029386.  
XX  
XX (PENN/) PENN S G.  
XX (RANK/) RANK D R.  
XX (HANZ/) HANZEL D K.  
XX  
XX Penn SG, Rank DR, Hanzel DK,  
XX WPI, 2004-119264/12.  
XX  
XX New human genome-derived single exon nucleic acid probes useful for human  
PT gene expression analysis, for identifying or characterizing alternative  
PT splicing events, for assessing genomic alterations or as tools for  
PT surveying tissues.  
XX  
XX Claim 45; SEQ ID NO 31343; 80pp; English.  
XX  
XX The invention relates to a nucleic acid probe for measuring human gene  
CC expression, comprising any of the 27,400 fully defined nucleotide  
CC sequences in the specification, or their complements or fragments, and  
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences  
CC fully defined in the specification. The probe is a single exon probe that  
CC hybridizes under high stringency conditions to a nucleic acid molecule  
CC expressed in human cells or tissues. Also included are a spatially-  
CC addressable set of single exon nucleic acid probes for measuring human  
CC gene expression (comprising a plurality of single exon nucleic acid  
CC probes cited above, where each of the plurality of probes is separately  
CC and addressably isolatable or amplifiable from the plurality), a single  
CC exon microarray for measuring human gene expression, a method of  
CC measuring human gene expression, a vector comprising the single exon  
CC probe cited above, an ORF-encoded peptide comprising at least 8  
CC contiguous amino acids of any of the above-mentioned amino acid  
CC sequences (optionally with conservative amino acid substitutions), an  
CC isolated antibody that binds specifically to a peptide cited above,  
CC a method of selling and/or licensing single exon probes or microarrays to  
CC a customer desiring to measure gene expression, a method of providing  
CC human gene expression data by subscrption, and a computer-readable  
CC storage medium which contains a database having a plurality of records  
CC (each record including data on the expression of a single exon probe  
CC cited above. The probe, methods and apparatus are useful in gene  
CC expression analysis. The probes may be used as tools for surveying  
CC tissues to detect the presence of expressed messages that contain their  
CC specific exon, or in constructing genome-derived single exon microarrays.  
CC In addition, the probes are used in identifying and characterizing  
CC alternative splicing events, in detecting and characterizing gross  
CC alterations in the genomic locus that includes their exon, in assessing  
CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
CC or in expressing the ORF-encoded peptide. The present sequence is a human  
CC single exon probe protein of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030194704  
XX  
XX Sequence 48 AA;  
SQ

Query Match 88.9%; Score 32; DB 8; Length 48;  
Best Local Similarity 100.0%; Pred. No. 5.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GPCV 6  
DB 39 GPCV 43  
RESULT 31  
AAM82764  
ID AAM82764 standard; protein, 53 AA.  
XX  
XX AAM82764;  
AC  
XX  
XX 07-NOV-2001 (first entry)  
DT  
XX  
XX Human immune/haematopoietic antigen SEQ ID NO:10357.  
DS  
XX  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis.  
XX  
XX Homo sapiens.  
XX  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US001354.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR  
XX 04-FEB-2000; 2000US-0180628P.  
PR  
XX 24-FEB-2000; 2000US-0184664P.  
PR  
XX 02-MAR-2000; 2000US-0186350P.  
PR  
XX 16-MAR-2000; 2000US-0189874P.  
PR  
XX 17-MAR-2000; 2000US-0190076P.  
PR  
XX 18-APR-2000; 2000US-0198123P.  
PR  
XX 19-MAY-2000; 2000US-0205515P.  
PR  
XX 07-JUN-2000; 2000US-0209467P.  
PR  
XX 28-JUN-2000; 2000US-0214886P.  
PR  
XX 30-JUN-2000; 2000US-0215135P.  
PR  
XX 07-JUL-2000; 2000US-0216647P.  
PR  
XX 07-JUL-2000; 2000US-0216880P.  
PR  
XX 11-JUL-2000; 2000US-0217487P.  
PR  
XX 11-JUL-2000; 2000US-0217496P.  
PR  
XX 14-JUL-2000; 2000US-0218290P.  
PR  
XX 26-JUL-2000; 2000US-0220963P.  
PR  
XX 26-JUL-2000; 2000US-0220964P.  
PR  
XX 14-AUG-2000; 2000US-0224518P.  
PR  
XX 14-AUG-2000; 2000US-0224519P.  
PR  
XX 14-AUG-2000; 2000US-0225213P.  
PR  
XX 14-AUG-2000; 2000US-0225214P.  
PR  
XX 14-AUG-2000; 2000US-0225265P.  
PR  
XX 14-AUG-2000; 2000US-0225267P.  
PR  
XX 14-AUG-2000; 2000US-0225268P.  
PR  
XX 14-AUG-2000; 2000US-0225270P.  
PR  
XX 14-AUG-2000; 2000US-0225447P.  
PR  
XX 14-AUG-2000; 2000US-0225757P.  
PR  
XX 14-AUG-2000; 2000US-0225758P.  
PR  
XX 14-AUG-2000; 2000US-0225759P.  
PR  
XX 18-AUG-2000; 2000US-0226811P.  
PR  
XX 22-AUG-2000; 2000US-0226868P.  
PR  
XX 22-AUG-2000; 2000US-0227183P.  
PR  
XX 23-AUG-2000; 2000US-0227009P.  
PR  
XX 30-AUG-2000; 2000US-0228924P.  
PR  
XX 01-SEP-2000; 2000US-0229287P.  
PR  
XX 01-SEP-2000; 2000US-0229343P.  
PR  
XX 01-SEP-2000; 2000US-0229344P.  
PR  
XX 01-SEP-2000; 2000US-0229345P.  
PR  
XX 09-SEP-2000; 2000US-0229309P.  
PR  
XX 05-SEP-2000; 2000US-0229513P.  
PR  
XX 06-SEP-2000; 2000US-0230437P.  
PR  
XX 08-SEP-2000; 2000US-0231242P.

PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0231415P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 25-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239933P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241825P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246473P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.

PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249298P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 01-DEC-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
DR N-PSDB; AAK55545.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
XX Claim 11; SEQ ID NO 10357; 3071pp + Sequence Listing; English.  
PS  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AAK62170 to AAK61921. (I) have cytosolic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patient's own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting the  
XX nucleic acids into a host cell and culturing the cell to express the  
XX protein. (I) proteins and polynucleotides may be used to prevent,  
XX diagnose and treat immune/hematopoietic-related diseases, especially  
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
XX to AAK67694 represent human immune/hematopoietic antigen genomic  
XX sequences from the present invention. AAK54942 to AAK54950 and AAK62169  
XX represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 53 AA;  
SQ  
Query Match 88.9%; Score 32; DB 4; Length 53;  
Best Local Similarity 100.0%; Pred. No. 5.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 AGGRC 5  
DB 39 AGGRC 43  
RESULT 32  
AAW05836  
ID AAW05836 standard; protein; 54 AA.  
XX  
XX AAW05836;  
XX  
XX 25-MAR-2003 (revised)  
DT 28-JAN-1997 (first entry)  
XX  
XX Mouse M-Serrate-1 (DSL domain upstream region).  
DB  
XX M-Serrate-1; Notch; cell differentiation; cell fate; tissue repair;  
KW central nervous system; cancer; therapy; diagnosis;  
KM

```

XX OS Mus sp.
XX PN W09627610-A1.
XX PD 12-SEP-1996.
XX PF 07-MAR-1996; 96WO-US003172.
XX PR 07-MAR-1995; 95US-00400159.
XX PA (UYTA ) UNIV YALE.
XX PA (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX PI Ishorowicz D, Henrique DMP, Lewis JH, Myat AM;
XX PI Artavanis-Tsakonas S, Mann RS, Gray GE;
XX DR WPI; 1996-425379/42.
XX DR N-PSDB; AA140095.
XX PT Vertebrate Serrate protein and related DNA - used to treat or prevent
XX PT malignancies characterized by increased Notch activity.
XX PS Disclousure; Page 65; 161pp; English.
XX CC A polypeptide (AAW05836) corresponds to an N-terminal sequence of mouse
XX CC Serrate homologue, M-Serrate-1, located upstream of the DSL domain. It is
XX CC the product of the 5' end (AA140095) of a cDNA clone isolated from mouse
XX CC 10.5-day embryo. Another polypeptide (AAW05837) located within the
XX CC epidermal growth factor-like repeats of Serrate was also identified. M-
XX CC Serrate-1 is expressed in the central nervous system, peripheral nervous
XX CC system, limb, kidney, lens and vascular system. It is a ligand for Notch
XX CC and is thought to play an important role in determining cell fate, esp.
XX CC in the central nervous system. (Updated on 25-MAR-2003 to correct PI
XX CC field.)
XX SQ Sequence 54 AA;
QY 1 AGGPG 5
DB 5 AGGPG 9
Query Match 88.9%; Score 32; DB 2; Length 54;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
RESULT 33
AAV59595
ID AAV59595 standard; protein; 54 AA.
XX AC AAV59595;
XX DT 05-APR-2000 (first entry)
XX DE Mouse Serrate protein sequence.
XX KW Serrate; cancerous condition; melanoma; lung cancer; breast cancer;
XX KW nervous system disorder; infection; nutritional disease; therapy;
XX KW cell proliferation promoter; tissue regeneration; mouse.
XX OS Mus sp.
XX PN US6004924-A.
XX PD 21-DEC-1999.
XX PF 06-MAR-1996; 96US-00611729.
XX PR 11-DEC-1991; 91US-00808458.
XX PR 14-SEP-1993; 93US-00121979.
XX PR 07-JUN-1994; 94US-00255102.
XX PR 07-MAR-1995; 95US-00400159.

```

```

XX PA (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX PA (UYTA ) UNIV YALE.
XX PI Ish-Horowicz D, Henrique DMP, Myat AM, Fleming RJ;
XX PI Artavanis-Tsakonas S, Gray GE, Mann RS, Lewis JH;
XX DR WPI; 2000-105089/09.
XX DR N-PSDB; AA249094.
XX PT Purified Serrate proteins useful for treating neoplasias, nervous
XX PT disorders and for promoting cell proliferation and tissue regeneration
XX PT and repair.
XX PS Claim 1, Col 135-136; 114pp; English.
XX CC This sequence represents a mouse serrate protein. The invention relates
XX CC to purified vertebrate (mouse, chick, and human) serrate proteins. The
XX CC Serrate proteins may be administered to treat a cancerous condition (e.g.
XX CC melanoma, lung cancer and breast cancer) by preventing progression from a
XX CC preneoplastic or nonmalignant state into a neoplastic or malignant state.
XX CC It may also be used to treat nervous system disorders (such as lesions
XX CC caused by infections, nutritional disease and toxic substances) and to
XX CC promote cell proliferation and tissue regeneration and repair. The
XX CC protein itself is administered to supplement a patient's own production
XX CC of Serrate proteins (if levels of expression are low) or to compensate
XX CC for expression of inactive proteins due to genetic mutations. The protein
XX CC may also be used in the production of antibodies against Serrate proteins
XX CC which may be used to either down regulate Serrate activity or to detect
XX CC Serrate proteins in samples (for example via enzyme-linked immunosorbent
XX CC assay (ELISA)). The proteins may also be used to study Serrate expression
XX CC and its role in metabolism and to assay for agents which modulate its
XX CC expression and activity
XX SQ Sequence 54 AA;
QY 1 AGGPG 5
DB 5 AGGPG 9
Query Match 88.9%; Score 32; DB 3; Length 54;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
RESULT 34
ADL71374
ID ADL71374 standard; peptide; 54 AA.
XX AC ADL71374;
XX DT 20-MAY-2004 (first entry)
XX DE Mouse serrate-1 partial protein #1.
XX KW Serrate protein; notch protein; breast cancer; fibrosarcoma; myxosarcoma;
XX KW nervous system disorder; motor neuron disorder; cirrhosis of liver; keloid; psoriasis;
XX KW benign dysproliferative disorder; carcinoma of liver; hepatocellular;
XX KW Chediak-Higashi syndrome; cytostatic; hepatocytic; vulnary;
XX KW antiproliferative; immunomodulator; serrate-1; mouse.
XX OS Mus sp.
XX PN US6703489-B1.
XX PD 09-MAR-2004.
XX PF 19-NOV-1998; 98US-00195524.
XX PR 07-MAR-1995; 95US-00400159.
XX PR 06-MAR-1996; 96US-00611729.
XX PA (UYTA ) UNIV YALE.

```



(IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.

XX Iah-Horowicz D, Henrique DMP, Lewis JH, Myat AM, Fleming RJ;  
 PI Attavanis-Tsakonas S, Mann RS, Gray GB;  
 XX WPI; 2004-224050/21.  
 DR N-PSDB; ADL71373.

XX New antibody binding to first Serrate protein or human Serrate protein,  
 PT useful for treating nervous system disorders, malignancies, cirrhosis of  
 PT liver.

XX Disclosure; SEQ ID NO 16; 117pp; English.

XX The present invention relates to novel serrate proteins which bind a  
 CC notch protein and polynucleotides encoding such proteins. Sequences of  
 CC the invention are useful for treating malignancies and related disorders  
 CC such as breast cancer, fibrosarcoma, myosarcoma, etc., nervous system  
 CC disorders, motor neuron disorders, benign dysplastic disorders,  
 CC cirrhosis of liver, keloid, psoriasis, Chediak-Higashi syndrome etc. They  
 CC are also useful for diagnosing diseases and disorders related to the  
 CC expression of Serrate protein. The present sequence is mouse serrate-1  
 CC partial protein used in the exemplification of the invention.

XX Sequence 54 AA;  
 SQ

Query Match 88.9%; Score 32; DB 8; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
 Db 5 AGGPC 9

RESULT 35  
 ADK34982  
 ID ADK34982 standard; protein; 71 AA.  
 XX  
 AC ADK34982;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Novel human polypeptide SegID7064.  
 XX  
 XX antiarthritic; antiparkinsonian; neuroprotective; noctropic;  
 KW immunosuppressive; cytostatic; antiproliferative; antiinflammatory;  
 KW antibacterial; antiviral; antifungal; antiparasitic; gene therapy;  
 KW arthritic; Parkinson's; Alzheimer's; autoimmune disease; cancer;  
 KW psoriasis; inflammatory bowel disease; infection; bacteria; virus;  
 KW fungus; parasite; human.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH Misc-difference 1..71  
 FT /label= OTHER  
 FT /note= "OTHER= All Xaa's in this sequence are unknown  
 FT amino acids or the site of a stop codon within the DNA  
 FT sequence"

XX WO200216439-A2.  
 EN  
 XX  
 XX 28-FEB-2002.  
 PD  
 XX  
 XX 05-MAR-2001; 2001WO-US004941.  
 PF  
 XX  
 XX 07-MAR-2000; 2000US-00519705.  
 PR  
 XX 19-MAY-2000; 2000US-00574454.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX  
 XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2002-280918/32.  
 DR  
 XX Isolated polynucleotide encoding bone marrow derived polypeptides useful  
 PT for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's  
 PT disease, and inflammatory bowel disease.  
 XX  
 XX Claim 20; SEQ ID NO 7064; 504pp; English.

XX This invention relates to a novel isolated polynucleotide comprising a  
 CC nucleotide sequence selected from one of 1680 sequences, a mature protein  
 CC coding portion of them, an active domain of them and their complementary  
 CC sequences. The invention may be useful for the production of compounds  
 CC with an antiarthritic, antiparkinsonian, neuroprotective, noctropic,  
 CC immunosuppressive, cytostatic, antiproliferative, antiinflammatory,  
 CC antibacterial, antiviral, antifungal or antiparasitic activity. In  
 CC addition, the disclosed sequences may be useful for gene therapy. The  
 CC polypeptides or their antibodies are useful for treating many diseases  
 CC such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,  
 CC psoriasis, inflammatory bowel disease and infections caused by bacteria,  
 CC viruses, fungi or parasites. The present sequence is that of a human  
 CC polypeptide of the invention.

XX Sequence 71 AA;  
 SQ

Query Match 88.9%; Score 32; DB 5; Length 71;  
 Best Local Similarity 100.0%; Pred. No. 7.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
 Db 28 GGPCV 32

RESULT 36  
 AAU65187  
 ID AAU65187 standard; protein; 74 AA.  
 XX  
 AC AAU65187;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #26083.  
 XX  
 XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 XX WO200181581-A2.  
 PN  
 XX  
 XX 01-NOV-2001.  
 PD  
 XX  
 XX 20-APR-2001; 2001WO-US012865.  
 PF  
 XX  
 XX 21-APR-2000; 2000US-0199047P.  
 PR  
 XX 02-JUN-2000; 2000US-020841P.  
 PR  
 XX 07-JUL-2000; 2000US-0216747P.  
 XX  
 XX (CORI-) CORIXA CORP.  
 PA  
 XX  
 XX Skelky YAM, Persing DH, Mitcham UT, Wang SS, Bhacia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 XX WPI; 2001-616774/71.  
 DR  
 XX N-PSDB; AAS59661.  
 DR  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris.  
 PT

PS Example 1, SEQ ID NO 26382, 1063pp; English.  
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SO Sequence 74 AA;  
Query Match 88.9%; Score 32; DB 4; Length 74;  
Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGGPC 5  
DB 29 AGGPC 33

RESULT 37  
ABG29902 standard; protein; 74 AA.  
XX  
XX ABG29902;  
AC  
XX  
XX 18-FEB-2002 (first entry)  
DT  
XX  
XX Novel human diagnostic protein #29893.  
DE  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
OS  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001WO-US008631.  
PE  
XX  
XX 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Dermanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS94089.  
XX  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 20; SEQ ID NO 60261; 103pp; English.  
PS  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC sequences. (I) is useful as hybridisation probe, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SO Sequence 74 AA;  
Query Match 88.9%; Score 32; DB 4; Length 74;  
Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGGPC 5  
DB 39 AGGPC 43

RESULT 38  
ABM61706 standard; protein; 74 AA.  
XX  
XX ABM61706;  
AC  
XX  
XX 20-OCT-2003 (first entry)  
DT  
XX  
XX Propionibacterium acnes predicted ORF-encoded polypeptide #26382.  
DE  
XX  
XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
KW immunostimulant; immune response; vaccine.  
OS  
XX  
XX Propionibacterium acnes.  
XX  
XX WO2003033515-A1.  
PN  
XX  
XX 24-APR-2003.  
PD  
XX  
XX 11-OCT-2002; 2002WO-US032727.  
PE  
XX  
XX 15-OCT-2001; 2001US-00978825.  
PR  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX  
XX Mitcham UL, Skeiky YAW, Persing DH, Bhattacha A, Maisonneuve JL;  
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
PI Barth B, Valiieve-Douglases J;  
PI  
XX  
XX WPI; 2003-381789/36.  
DR N-PSDB; ACF64590.  
XX  
XX  
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a P. acnes protein.  
XX  
XX Example 1, SEQ ID NO 26382, 1481pp; English.  
PS  
XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC encoding a Propionibacterium acnes protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to

CC immunogenic fragments of *P. acnes* polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a *P. acnes*  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition (comprising *P. acnes* polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of *P. acnes* in a  
 CC patient; and a method for inhibiting the development of *P. acnes* in a  
 CC patient. The *P. acnes* polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating *acne*  
 CC vulgaris, or for stimulating an immune response specific for a *P. acnes*  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridization. The vaccine composition is useful for the  
 CC stimulation of an immune response against *P. acnes*, or for treating *acne*,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
 CC reading frame) contained within the *P. acnes* polynucleotides of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

SO Sequence 74 AA:

Query Match 88.9%; Score 32; DB 6; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
 |||||  
 Db 29 AGGPC 33

RESULT 39  
 ADA57035  
 ID ADA57035 standard; protein; 88 AA.  
 AC ADA57035;  
 XX 20-NOV-2003 (first entry)  
 DT  
 XX Human secreted protein #318.  
 DE  
 XX Immunosuppressive; antiinflammatory; antiaesthetic; antiallergic;  
 KM cytotoxic; cerebroprotective; neuroprotective; neurotropic;  
 KM cardiovascular; antiarteriosclerotic; gene therapy;  
 KM human secreted protein; immune disorder; inflammation;  
 KM respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;  
 KM inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;  
 KM multiple sclerosis; ischaemic brain injury; Parkinson's disease;  
 KM Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;  
 KM triple helix formation; antisense gene therapy; forensic biology.

OS Homo sapiens.  
 OS  
 XX  
 PN WO2002102994-A2.  
 XX  
 PD 27-DEC-2002.  
 XX  
 XX 19-MAR-2002; 2002WC-US008278.  
 XX  
 XX 21-MAR-2001; 2001US-0277340P.  
 PR 19-JUL-2001; 2001US-0306171P.  
 PR 13-NOV-2001; 2001US-0331287P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Rosen CA, Ruben SM,  
 PI  
 XX WPI, 2003-167512/16.  
 DR

DR N-PSDB; ADA56139.  
 XX  
 XX New human secreted polypeptides and polynucleotides, useful for  
 PT diagnosing, treating or preventing e.g. immune disorders, inflammatory  
 PT neurodegenerative disorders, cancers, CNS disorders, or  
 PT neurodegenerative disorders.

PS Claim 13; SEQ ID NO 1225; 1754bp; English.

XX The invention relates to 592 new human secreted polypeptides useful for  
 CC diagnosing, treating or preventing e.g. immune disorders, inflammatory  
 CC conditions, respiratory disorders, cancers, CNS disorders, or  
 CC neurodegenerative disorders, or polypeptides comprising an amino acid  
 CC sequence at least 95% identical to the new sequences. The polypeptides,  
 CC antibodies or antibody fragments that bind to the polypeptides, nucleic  
 CC acids encoding the polypeptides, agonists or antagonists that bind to  
 CC the polypeptide, are useful in preparing diagnostic or pharmaceutical  
 CC compositions for diagnosing, treating or preventing an e.g. immune  
 CC disorder, inflammatory conditions (e.g. inflammatory bowel disease,  
 CC nephritis or Crohn's disease), respiratory disorders (e.g. asthma and  
 CC allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders  
 CC (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative  
 CC disorders (e.g. Parkinson's disease or Alzheimer's disease), and  
 CC cardiovascular disorders (e.g. atherosclerosis or myocarditis). The  
 CC polynucleotides are useful for chromosome identification, chromosome  
 CC mapping, for controlling gene expression through triple helix formation  
 CC or antisense DNA or RNA, in gene therapy, for identifying individuals  
 CC from minute biological samples, in forensic biology, and as hybridization  
 CC probes. The polypeptides are useful for as molecular weight markers on  
 CC sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)  
 CC gels, to raise antibodies, for testing biological activities, and for  
 CC treating or preventing neural disorders, immune system disorders,  
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,  
 CC renal, proliferative and/or cancerous diseases. This sequence corresponds  
 CC to one of the polypeptide of the invention. Note: The sequence data for  
 CC this patent did form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 SO Sequence 88 AA:

Query Match 88.9%; Score 32; DB 6; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
 |||||  
 Db 58 AGGPC 62

RESULT 40  
 ADA40890  
 ID ADA40890 standard; protein; 88 AA.  
 AC ADA40890;  
 XX 20-NOV-2003 (first entry)  
 DT  
 XX Human secreted protein.  
 DE  
 XX Human, secreted protein; cancer; hyperproliferative disorder;  
 KM rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;  
 KM anaemia; allergic reaction; asthma; cardiovascular disorder;  
 KM wound healing; cytotoxic; immunosuppressive; neuroprotective;  
 KM antiviral; antiinflammatory; hepatotropic; antidiabetic; antiinflammatory;  
 KM vulnery; cardiac; gene therapy.

OS Homo sapiens.  
 OS  
 XX  
 PN WO2002102993-A2.  
 XX  
 PD 27-DEC-2002.  
 XX

PF 19-MAR-2002; 2002WO-US008123.  
XX  
XX 21-MAR-2001; 2001US-0277340P.  
PR 19-JUL-2001; 2001US-0306171P.  
PR 13-NOV-2001; 2001US-0331287P.  
XX  
XX (HDMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM;  
XX  
XX WPI; 2003-175238/17.  
DR

XX New human secreted proteins and nucleic acid molecules, useful for  
PT preparing a diagnostic or pharmaceutical composition for diagnosing,  
PT preventing or treating cancer or other hyperproliferative disorder,  
PT asthma, allergies or AIDS.  
XX  
XX

PS Claim 1; SEQ ID NO 1272; 3205pp; English.

XX The invention relates to novel genes ADA39629-ADA40565 and proteins  
CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,  
CC treating or ameliorating medical conditions e.g. by protein or gene  
CC therapy. The polypeptides, nucleic acid molecules, antibodies or their  
CC fragments, and agonists or antagonists that bind to the polypeptide are  
CC useful for preparing a diagnostic or pharmaceutical composition for  
CC diagnosing or treating cancer or other hyperproliferative disorder. The  
CC polypeptides and nucleic acid molecules are also useful for detecting,  
CC preventing, diagnosing, prognosticating, treating or ameliorating cancer  
CC or other hyperproliferative disorders including neoplasms, autoimmune  
CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus  
CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic  
CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,  
CC thrombocytopenia), allergic reactions including asthma or eczema,  
CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory  
CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders  
CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,  
CC fungal or viral infections including HIV/AIDS), or wound healing and  
CC disorders of epithelial cell proliferation. The nucleic acids are also  
CC useful for chromosome identification, radiation hybrid mapping or long-  
CC range restriction mapping, as molecular weight markers, or as  
CC hybridization or diagnostic probes. The polypeptides and antibodies are  
CC useful for providing immunological probes for differential identification  
CC of the tissues immunohistochemistry assays. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 88 AA;

Query Match 88.9% Score 32; DB 6; Length 88;  
Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
|||  
Db 58 AGGPC 62

Search completed: December 30, 2004, 15:08:57  
Job time : 10.7879 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 30, 2004, 15:15:03 ; Search time 156 Seconds

(without alignments)  
32.194 Million cell updates/sec

Title: US-10-719-385-20

Perfect score: 14

Sequence: 1 MIRSKITSVLSFC 14

Scoring table: Gapex 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 150 summaries

Database : A\_Geneseq\_23Sep04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query length	DB ID	Description
1	14	100.0	14	8	ADO71787 BFLP0169
2	14	100.0	1752	8	ADO71766 BFLP0169
3	14	100.0	1753	8	ADO71777 BFLP0169
4	14	100.0	1753	8	ADO71767 BFLP0169
5	14	100.0	1753	8	ADO71775 BFLP0169
6	14	100.0	1753	8	ADO71768 BFLP0169
7	14	100.0	1753	8	ADO71771 BFLP0169
8	14	100.0	1753	8	ADO71780 BFLP0169
9	14	100.0	1753	8	ADO71772 BFLP0169
10	14	100.0	1753	8	ADO71764 BFLP0169
11	14	100.0	1753	8	ADO71765 BFLP0169
12	14	100.0	1753	8	ADO71773 BFLP0169
13	14	100.0	1753	8	ADO71778 BFLP0169
14	14	100.0	1753	8	ADO71769 BFLP0169
15	14	100.0	1753	8	ADO71770 BFLP0169
16	14	100.0	1753	8	ADO71774 BFLP0169
17	14	100.0	1753	8	ADO71779 BFLP0169
18	14	100.0	1753	8	ADO71763 BFLP0169
19	14	100.0	1753	8	ADO71776 BFLP0169
20	14	100.0	1753	8	ADO71777 BFLP0169
21	14	100.0	1753	8	ADO71769 BFLP0169
22	14	100.0	1753	8	ADO71770 BFLP0169
23	14	100.0	1753	8	ADO71774 BFLP0169
24	14	100.0	1753	8	ADO71779 BFLP0169
25	14	100.0	1753	8	ADO71763 BFLP0169
26	14	100.0	1753	8	ADO71776 BFLP0169
27	14	100.0	1753	8	ADO71777 BFLP0169
28	14	100.0	1753	8	ADO71769 BFLP0169
29	14	100.0	1753	8	ADO71770 BFLP0169
30	14	100.0	1753	8	ADO71774 BFLP0169
31	14	100.0	1753	8	ADO71779 BFLP0169
32	14	100.0	1753	8	ADO71763 BFLP0169
33	14	100.0	1753	8	ADO71776 BFLP0169
34	14	100.0	1753	8	ADO71777 BFLP0169
35	14	100.0	1753	8	ADO71769 BFLP0169
36	14	100.0	1753	8	ADO71770 BFLP0169
37	14	100.0	1753	8	ADO71774 BFLP0169
38	14	100.0	1753	8	ADO71779 BFLP0169
39	14	100.0	1753	8	ADO71763 BFLP0169
40	14	100.0	1753	8	ADO71776 BFLP0169
41	14	100.0	1753	8	ADO71777 BFLP0169
42	14	100.0	1753	8	ADO71769 BFLP0169
43	14	100.0	1753	8	ADO71770 BFLP0169
44	14	100.0	1753	8	ADO71774 BFLP0169
45	14	100.0	1753	8	ADO71779 BFLP0169
46	14	100.0	1753	8	ADO71763 BFLP0169
47	14	100.0	1753	8	ADO71776 BFLP0169
48	14	100.0	1753	8	ADO71777 BFLP0169
49	14	100.0	1753	8	ADO71769 BFLP0169
50	14	100.0	1753	8	ADO71770 BFLP0169
51	14	100.0	1753	8	ADO71774 BFLP0169
52	14	100.0	1753	8	ADO71779 BFLP0169
53	14	100.0	1753	8	ADO71763 BFLP0169
54	14	100.0	1753	8	ADO71776 BFLP0169
55	14	100.0	1753	8	ADO71777 BFLP0169
56	14	100.0	1753	8	ADO71769 BFLP0169
57	14	100.0	1753	8	ADO71770 BFLP0169
58	14	100.0	1753	8	ADO71774 BFLP0169
59	14	100.0	1753	8	ADO71779 BFLP0169
60	14	100.0	1753	8	ADO71763 BFLP0169
61	14	100.0	1753	8	ADO71776 BFLP0169
62	14	100.0	1753	8	ADO71777 BFLP0169
63	14	100.0	1753	8	ADO71769 BFLP0169
64	14	100.0	1753	8	ADO71770 BFLP0169
65	14	100.0	1753	8	ADO71774 BFLP0169
66	14	100.0	1753	8	ADO71779 BFLP0169
67	14	100.0	1753	8	ADO71763 BFLP0169
68	14	100.0	1753	8	ADO71776 BFLP0169
69	14	100.0	1753	8	ADO71777 BFLP0169
70	14	100.0	1753	8	ADO71769 BFLP0169
71	14	100.0	1753	8	ADO71770 BFLP0169
72	14	100.0	1753	8	ADO71774 BFLP0169
73	14	100.0	1753	8	ADO71779 BFLP0169
74	14	100.0	1753	8	ADO71763 BFLP0169
75	14	100.0	1753	8	ADO71776 BFLP0169
76	14	100.0	1753	8	ADO71777 BFLP0169
77	14	100.0	1753	8	ADO71769 BFLP0169
78	14	100.0	1753	8	ADO71770 BFLP0169
79	14	100.0	1753	8	ADO71774 BFLP0169
80	14	100.0	1753	8	ADO71779 BFLP0169
81	14	100.0	1753	8	ADO71763 BFLP0169
82	14	100.0	1753	8	ADO71776 BFLP0169
83	14	100.0	1753	8	ADO71777 BFLP0169
84	14	100.0	1753	8	ADO71769 BFLP0169
85	14	100.0	1753	8	ADO71770 BFLP0169
86	14	100.0	1753	8	ADO71774 BFLP0169
87	14	100.0	1753	8	ADO71779 BFLP0169
88	14	100.0	1753	8	ADO71763 BFLP0169
89	14	100.0	1753	8	ADO71776 BFLP0169
90	14	100.0	1753	8	ADO71777 BFLP0169
91	14	100.0	1753	8	ADO71769 BFLP0169
92	14	100.0	1753	8	ADO71770 BFLP0169
93	14	100.0	1753	8	ADO71774 BFLP0169
94	14	100.0	1753	8	ADO71779 BFLP0169
95	14	100.0	1753	8	ADO71763 BFLP0169
96	14	100.0	1753	8	ADO71776 BFLP0169
97	14	100.0	1753	8	ADO71777 BFLP0169
98	14	100.0	1753	8	ADO71769 BFLP0169
99	14	100.0	1753	8	ADO71770 BFLP0169
100	14	100.0	1753	8	ADO71774 BFLP0169

26	7	50.0	431	8	ADO25474
27	7	50.0	1019	4	AA92745
28	7	50.0	1019	4	AB87944
29	7	50.0	1019	4	ADL65625
30	6	42.9	12	4	AA55995
31	6	42.9	12	4	ADA88830
32	6	42.9	17	5	ABG78459
33	6	42.9	77	3	AA600649
34	6	42.9	85	4	ABG06877
35	6	42.9	90	6	AA939937
36	6	42.9	90	6	ABM36456
37	6	42.9	130	7	ABU17550
38	6	42.9	133	7	ADC94908
39	6	42.9	135	7	ADF04124
40	6	42.9	139	5	ABP10237
41	6	42.9	143	7	ADC88364
42	6	42.9	183	7	ADF06790
43	6	42.9	184	8	ADO03923
44	6	42.9	186	8	ADO03922
45	6	42.9	192	5	ADK34600
46	6	42.9	194	5	AA93258
47	6	42.9	204	4	ADL30674
48	6	42.9	207	4	AAU19352
49	6	42.9	216	3	AA616571
50	6	42.9	218	5	ABP27242
51	6	42.9	238	8	ADK47075
52	6	42.9	242	3	AA628834
53	6	42.9	242	3	AA635588
54	6	42.9	270	7	ADC94107
55	6	42.9	284	3	AA606151
56	6	42.9	289	3	AA627262
57	6	42.9	289	3	AA640485
58	6	42.9	289	5	AA625109
59	6	42.9	289	5	AAU72581
60	6	42.9	289	8	ADN72347
61	6	42.9	298	8	ADN61913
62	6	42.9	302	4	AA604555
63	6	42.9	308	6	ADK50458
64	6	42.9	311	4	AA672083
65	6	42.9	311	4	AAU24523
66	6	42.9	311	5	AAU85862
67	6	42.9	311	5	AAU85863
68	6	42.9	311	5	ABP95796
69	6	42.9	311	5	AAU85147
70	6	42.9	311	5	ABO63269
71	6	42.9	314	5	ABP51568
72	6	42.9	314	7	ADL04261
73	6	42.9	316	6	ABU19105
74	6	42.9	317	8	ADK36453
75	6	42.9	317	8	ADU19375
76	6	42.9	321	5	ABP90173
77	6	42.9	321	5	ADU19416
78	6	42.9	322	5	ABP28189
79	6	42.9	322	5	AA606150
80	6	42.9	326	6	ABP56928
81	6	42.9	326	6	AAU98390
82	6	42.9	327	3	AA627261
83	6	42.9	329	4	AAU89960
84	6	42.9	332	6	ABU18942
85	6	42.9	332	6	ABW72372
86	6	42.9	336	5	AA606149
87	6	42.9	340	5	ABP99397
88	6	42.9	344	7	ADC95919
89	6	42.9	360	3	AAU77470
90	6	42.9	360	5	ABU61835
91	6	42.9	374	5	AAU80496
92	6	42.9	376	5	ABP30123
93	6	42.9	384	7	ADK08050
94	6	42.9	387	7	ADJ70426
95	6	42.9	403	5	ABG31768
96	6	42.9	404	6	ABU02498
97	6	42.9	404	6	ABP81646
98	6	42.9	413	8	ADL66350

ADQ5474	E. faecali
AA92745	C. glutami
AB87944	Coryneb
ADL65625	C. glutam
AA55995	Internal
ADA88830	Internal
ABG78459	H. RRS11
AA600649	Human sec
ABG06877	Novel hum
AA939937	Propionib
ABM36456	Propionib
ABU17550	Protein e
ADC94908	E. faecali
ADF04124	Bacterial
ABP10237	Human ORP
ADC88364	Ribosomal
ADF06790	Bacterial
ADO03923	Rabbit fu
ADO03922	Rat full
ADK34600	Novel hum
AA93258	Human pol
ADL30674	Human pro
AAU19352	Human G p
AA616571	Arabiidops
ABP27242	Streptoco
ADK47075	Streptoco
AA628834	Arabiidops
AA635588	Arabiidops
ADC94107	E. faecali
AA606151	Arabiidops
ABP27262	Arabiidops
AA640485	Arabiidops
AA625109	Arabiidops
AAU72581	Arabiidops
ADN72347	

99	6	42.9	413	8	ADN9857	Adn9857 Novel hum
100	6	42.9	416	5	ABP27243	Abp27243 Streptoco
101	6	42.9	429	3	AGS3587	AgS3587 Arabidops
102	6	42.9	429	3	AGS28833	AgS28833 Arabidops
103	6	42.9	430	6	ABP75913	Abp75913 Human sec
104	6	42.9	431	6	ABU29849	Abu29849 Protein e
105	6	42.9	433	3	AGS28832	AgS28832 Arabidops
106	6	42.9	433	3	AGS3586	AgS3586 Arabidops
107	6	42.9	438	8	ADN9854	Adn9854 Novel hum
108	6	42.9	447	8	ADP49384	Adp49384 Xanthomon
109	6	42.9	468	7	ADM27042	Adm27042 Hyperther
110	6	42.9	482	3	AGS1506	AgS1506 Arabidops
111	6	42.9	482	5	ABB90925	Abb90925 Herbi:cida
112	6	42.9	494	2	AAW30711	Aaw30711 Human ubi
113	6	42.9	496	8	ADP47895	Adp47895 Human CAT
114	6	42.9	497	7	ADC31203	Adc31203 Human nov
115	6	42.9	497	7	ADC27031	Adc27031 Human deu
116	6	42.9	501	3	AGA47662	AgA47662 Arabidops
117	6	42.9	519	4	ABG12217	Abg12217 Novel hum
118	6	42.9	519	4	ABG05882	Abg05882 Novel hum
119	6	42.9	527	3	AGS3833	AgS3833 Arabidops
120	6	42.9	530	3	AAW747471	Aaw747471 Human deu
121	6	42.9	530	6	AAAG64049	AAAG64049 Human deu
122	6	42.9	530	6	ABU61836	Abu61836 Updated h
123	6	42.9	530	6	ABU10207	Abu10207 Human ubi
124	6	42.9	530	6	ABR43250	AbR43250 Human PMW
125	6	42.9	530	7	ADC27021	Adc27021 Human deu
126	6	42.9	530	7	ADC27033	Adc27033 Human deu
127	6	42.9	530	7	ADC27019	Adc27019 Human deu
128	6	42.9	530	7	ADC27037	Adc27037 Human deu
129	6	42.9	530	7	ADC27039	Adc27039 Human deu
130	6	42.9	530	7	ADC27009	Adc27009 Human deu
131	6	42.9	530	7	ADC27035	Adc27035 Human deu
132	6	42.9	530	7	ADC27015	Adc27015 Human deu
133	6	42.9	530	8	ADN9856	Adn9856 Novel hum
134	6	42.9	538	4	ABG07586	Abg07586 Novel hum
135	6	42.9	539	4	AAW31604	Aaw31604 Aridno act
136	6	42.9	540	2	AAW07871	Aaw07871 GDV (or G
137	6	42.9	558	6	ABU25727	Abu25727 Protein e
138	6	42.9	565	7	ADC27023	Adc27023 Human deu
139	6	42.9	566	6	ABR53129	AbR53129 Protein s
140	6	42.9	566	7	ADK62560	AdK62560 Disease c
141	6	42.9	574	7	ADC27017	Adc27017 Human deu
142	6	42.9	578	2	AAW09065	Aaw09065 Human com
143	6	42.9	584	8	ADH17324	Adh17324 Human NOV
144	6	42.9	584	8	ABO58812	AbO58812 Human gen
145	6	42.9	587	6	ABR53423	AbR53423 Protein s
146	6	42.9	587	4	ADK64684	AdK64684 Disease c
147	6	42.9	597	4	ABR11735	AbR11735 Human vas
148	6	42.9	597	7	ADBO3343	AdBO3343 Novel pro
149	6	42.9	597	8	ADM87677	AdM87677 Human EST
150	6	42.9	615	3	AAW35832	Aaw35832 Arabidops

## ALIGNMENTS

## RESULT 1

ADO71787  
ID ADO71787 standard; protein; 14 AA.

ADO71787;

26-AUG-2004 (first entry)

BFLP0169 protein fragment, SEQ ID 20.

Antiinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
Kidney; lupus nephritis; autoimmune disorder; lupus.

Homo sapiens.

WO2004048521-A2.

XX 10-JUN-2004.  
PD 21-NOV-2003; 2003WO-US037339.  
XX 21-NOV-2002; 2002US-0428094P.  
XX (AMHP) WYETH.  
XX O'Loole M, Mounts WM, Shojaae N;  
XX WPI; 2004-441163/41.  
XX  
XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
PS Claim 20; Page; 83pp; English.  
XX  
XX The present invention relates to human BFLP0169 (ADO71763) and its coding  
CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
CC for diagnosing and treating autoimmune disorder, preferably lupus or  
CC lupus nephritis. They are also useful in the generation of antibodies  
CC that bind immunospecifically to the nucleic acids. The present sequence  
CC was used to illustrate the invention. Note: The present sequence was not  
CC shown in the specification, but was derived from information given on  
CC page 13 and ADO71763.

Sequence 14 AA;

Query Match 100.0%; Score 14; DB 8; Length 14;  
Best Local Similarity 100.0%; Pred. No. 6; 1e-08;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
Db 1 MIRSKITSVLSFC 14

## RESULT 2

ADO71766  
ID ADO71766 standard; protein; 1752 AA.

ADO71766;

26-AUG-2004 (first entry)

BFLP0169 variant R267K, SEQ ID 5.

Antiinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
Kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutacin.

Homo sapiens.

Synthetic.

Key Location/Qualifiers

FT Misc-difference 267 /note= "Wild-type R replaced with K"

WO2004048521-A2.

10-JUN-2004.

21-NOV-2003; 2003WO-US037339.

21-NOV-2002; 2002US-0428094P.

(AMHP) WYETH.

O'Loole M, Mounts WM, Shojaae N;

WPI; 2004-441163/41.

PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
XX  
XX  
PS Example 4; Page 64-65; 83pp; English.

CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
CC for diagnosing and treating autoimmune disorder, preferably lupus or  
CC lupus nephritis. They are also useful in the generation of antibodies  
CC that bind immunospecifically to the nucleic acids. The present sequence  
CC is a BFLP0169 variant.  
XX  
SQ Sequence 1752 AA;

Query Match 100.0%; Score 14; DB 8; Length 1752;  
Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRSKITSVLSFC 14  
Db 1 MIRSKITSVLSFC 14

RESULT 3  
ADO71777  
ID ADO71777 standard; protein; 1753 AA.  
XX  
AC ADO71777;  
XX

26-AUG-2004 (first entry)  
XX

BFLP0169 variant K1331R, SEQ ID 16.  
XX

Antiinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.  
XX

OS Homo sapiens.  
OS Synthetic.

Key Location/Qualifiers  
FT Misc-difference 1331  
FT /note= "Wild-type K replaced with R"

MO2004048521-A2.  
XX

10-JUN-2004.  
XX

21-NOV-2003; 2003WO-US037339.  
XX

21-NOV-2002; 2002US-0428094P.  
XX

(AMHP ) WYETH.  
XX

O'coole M, Mounts WM, Shojaaee N;  
XX

WPI; 2004-441163/41.  
XX

New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
XX

Example 15; Page 71; 83pp; English.  
XX

CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
CC for diagnosing and treating autoimmune disorder, preferably lupus or  
CC lupus nephritis. They are also useful in the generation of antibodies  
CC that bind immunospecifically to the nucleic acids. The present sequence  
CC is a BFLP0169 variant.  
XX  
SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRSKITSVLSFC 14  
Db 1 MIRSKITSVLSFC 14

RESULT 4  
ADO71767  
ID ADO71767 standard; protein; 1753 AA.  
XX  
AC ADO71767;  
XX

26-AUG-2004 (first entry)  
XX

BFLP0169 variant Q355N, SEQ ID 6.  
XX

Antiinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.  
XX

OS Homo sapiens.  
OS Synthetic.

Key Location/Qualifiers  
FT Misc-difference 355  
FT /note= "Wild-type Q replaced with N"

MO2004048521-A2.  
XX

10-JUN-2004.  
XX

21-NOV-2003; 2003WO-US037339.  
XX

21-NOV-2002; 2002US-0428094P.  
XX

(AMHP ) WYETH.  
XX

O'coole M, Mounts WM, Shojaaee N;  
XX

WPI; 2004-441163/41.  
XX

New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
XX

Example 5; Page 65; 83pp; English.  
XX

CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
CC for diagnosing and treating autoimmune disorder, preferably lupus or  
CC lupus nephritis. They are also useful in the generation of antibodies  
CC that bind immunospecifically to the nucleic acids. The present sequence  
CC is a BFLP0169 variant.  
XX  
SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRSKITSVLSFC 14  
Db 1 MIRSKITSVLSFC 14

RESULT 5  
ADO71775  
ID ADO71775 standard; protein; 1753 AA.  
XX  
AC ADO71775;  
XX

DT 26-AUG-2004 (first entry)  
 XX BFLP0169 variant A11528, SEQ ID 14.  
 DE BFLP0169 variant A11528, SEQ ID 14.  
 XX Anti-inflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutcin.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1152 /note= "Wild-type A replaced with S"  
 PT  
 XX WO2004048521-A2.  
 PN  
 XX 10-JUN-2004.  
 PD  
 XX 21-NOV-2003; 2003WO-US037339.  
 PF  
 XX 21-NOV-2002; 2002US-0428094P.  
 PR  
 XX (AMHP ) WYETH.  
 PA  
 XX O'coole M, Mounts WM, Shojaee N;  
 PI  
 XX WPI; 2004-441163/41.  
 DR  
 XX  
 XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 PS  
 XX Example 13; Page 70; 83pp; English.  
 XX  
 CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.  
 CC  
 XX  
 SQ Sequence 1753 AA;  
 Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MIRSKITSVLSFC 14  
 DB 1 MIRSKITSVLSFC 14.  
 RESULT 6  
 ADO71768  
 ID ADO71768 standard; protein; 1753 AA.  
 XX  
 AC ADO71768;  
 XX  
 DT 26-AUG-2004 (first entry)  
 DE BFLP0169 variant H446R, SEQ ID 7.  
 XX  
 XX Anti-inflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutcin.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 446 /note= "Wild-type H replaced with R"  
 PT  
 XX WO2004048521-A2.  
 PN

XX 10-JUN-2004.  
 PD  
 XX 21-NOV-2003; 2003WO-US037339.  
 PF  
 XX 21-NOV-2002; 2002US-0428094P.  
 PR  
 XX (AMHP ) WYETH.  
 PA  
 XX O'coole M, Mounts WM, Shojaee N;  
 PI  
 XX WPI; 2004-441163/41.  
 DR  
 XX  
 XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 PS  
 XX Example 6; Page 66; 83pp; English.  
 XX  
 CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.  
 CC  
 XX  
 SQ Sequence 1753 AA;  
 Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MIRSKITSVLSFC 14  
 DB 1 MIRSKITSVLSFC 14  
 RESULT 7  
 ADO71771  
 ID ADO71771 standard; protein; 1753 AA.  
 XX  
 AC ADO71771;  
 XX  
 DT 26-AUG-2004 (first entry)  
 DE BFLP0169 variant N749D, SEQ ID 10.  
 XX  
 XX Anti-inflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutcin.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 749 /note= "Wild-type N replaced with D"  
 PT  
 XX WO2004048521-A2.  
 PN  
 XX 10-JUN-2004.  
 PD  
 XX 21-NOV-2003; 2003WO-US037339.  
 PF  
 XX 21-NOV-2002; 2002US-0428094P.  
 PR  
 XX (AMHP ) WYETH.  
 PA  
 XX O'coole M, Mounts WM, Shojaee N;  
 PI  
 XX WPI; 2004-441163/41.  
 DR  
 XX  
 XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.



XX Example 9, Page 67-68; 83bp; English.

CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.

XX Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;

Best Local Similarity 100.0%; Pred. No. 3.9e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

AC ADO71780;

DT 26-AUG-2004 (first entry)

DE BFLP0169 variant F1706H, SEQ ID 19.

XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KW kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.

XX Homo sapiens.  
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 1706 /note= "Wild-type F replaced with H"

FN WO2004048521-A2.

PD 10-JUN-2004.

PP 21-NOV-2003; 2003WO-US037339.

PR 21-NOV-2002; 2002US-0428094P.

PA (AMHP ) WYETH.

PI O'coole M, Mounts WM, Shojaaee N;

DR WPI; 2004-441163/41.

XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.

PS Example 18; Page 73; 83bp; English.

CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.

XX Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;

Best Local Similarity 100.0%; Pred. No. 3.9e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

RESULT 9  
 ADO71772  
 ID ADO71772 standard; protein; 1753 AA.

AC ADO71772;

DT 26-AUG-2004 (first entry)

DE BFLP0169 variant V838M, SEQ ID 11.

XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KW kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.

XX Homo sapiens.  
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 838 /note= "Wild-type V replaced with M"

FN WO2004048521-A2.

PD 10-JUN-2004.

PP 21-NOV-2003; 2003WO-US037339.

PR 21-NOV-2002; 2002US-0428094P.

PA (AMHP ) WYETH.

PI O'coole M, Mounts WM, Shojaaee N;

DR WPI; 2004-441163/41.

XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.

PS Example 10; Page 68; 83bp; English.

CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.

XX Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;

Best Local Similarity 100.0%; Pred. No. 3.9e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

RESULT 10

ADO71764  
 ID ADO71764 standard; protein; 1753 AA.

AC ADO71764;

DT 26-AUG-2004 (first entry)

```

DE BFLP0169 variant L78V, SEQ ID 3.
XX
XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;
KW kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.
XX
XX Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 78 /note= "Wild-type L replaced with V"
XX
XX WO2004048521-A2.
XX
XX 10-JUN-2004.
XX
XX 21-NOV-2003; 2003WO-US037339.
XX
XX 21-NOV-2002; 2002US-0428094P.
XX
XX (AMHP ) WYETH.
XX
XX O'coole M, Mounts WM, Shojasee N;
XX
XX WPI; 2004-441163/41.
XX
XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing
XX or treating autoimmune disorder, e.g. lupus or lupus nephritis.
XX
XX Example 2, Page 63-64, 83pp; English.
XX
XX The present invention relates to human BFLP0169 (ADO71763) and its coding
XX sequence (ADO71762). BFLP0169 gene expression is increased in kidney
XX tissue in mice with lupus nephritis. The BFLP0169 sequences are useful
XX for diagnosing and treating autoimmune disorder, preferably lupus or
XX lupus nephritis. They are also useful in the generation of antibodies
XX that bind immunospecifically to the nucleic acids. The present sequence
XX is a BFLP0169 variant.
XX
XX Sequence 1753 AA;
XX
XX Query Match 100.0%; Score 14; DB 8; Length 1753;
XX Best Local Similarity 100.0%; Pred. No. 3.9e-06;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MIRSKITSVLSFC 14
XX 1 MIRSKITSVLSFC 14
XX
XX Db
XX
XX RESULT 11
XX ADO71765
XX ID ADO71765 standard; protein; 1753 AA.
XX
XX ADO71765;
XX
XX 26-AUG-2004 (first entry)
XX
XX BFLP0169 variant L198I, SEQ ID 4.
XX
XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;
KW kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX Key Location/Qualifiers
XX FH Misc-difference 198 /note= "Wild-type L replaced with I"
XX FT or treating autoimmune disorder, e.g. lupus or lupus nephritis.
XX
XX WO2004048521-A2.
XX
XX 10-JUN-2004.
XX

```

```

XX
XX 21-NOV-2003; 2003WO-US037339.
XX
XX 21-NOV-2002; 2002US-0428094P.
XX
XX (AMHP ) WYETH.
XX
XX O'coole M, Mounts WM, Shojasee N;
XX
XX WPI; 2004-441163/41.
XX
XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing
XX or treating autoimmune disorder, e.g. lupus or lupus nephritis.
XX
XX Example 3, Page 64, 83pp; English.
XX
XX The present invention relates to human BFLP0169 (ADO71763) and its coding
XX sequence (ADO71762). BFLP0169 gene expression is increased in kidney
XX tissue in mice with lupus nephritis. The BFLP0169 sequences are useful
XX for diagnosing and treating autoimmune disorder, preferably lupus or
XX lupus nephritis. They are also useful in the generation of antibodies
XX that bind immunospecifically to the nucleic acids. The present sequence
XX is a BFLP0169 variant.
XX
XX Sequence 1753 AA;
XX
XX Query Match 100.0%; Score 14; DB 8; Length 1753;
XX Best Local Similarity 100.0%; Pred. No. 3.9e-06;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MIRSKITSVLSFC 14
XX 1 MIRSKITSVLSFC 14
XX
XX Db
XX
XX RESULT 12
XX ADO71773
XX ID ADO71773 standard; protein; 1753 AA.
XX
XX ADO71773;
XX
XX 26-AUG-2004 (first entry)
XX
XX BFLP0169 variant G958T, SEQ ID 12.
XX
XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;
KW kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX Key Location/Qualifiers
XX FH Misc-difference 958 /note= "Wild-type G replaced with T"
XX FT or treating autoimmune disorder, e.g. lupus or lupus nephritis.
XX
XX WO2004048521-A2.
XX
XX 10-JUN-2004.
XX
XX 21-NOV-2003; 2003WO-US037339.
XX
XX 21-NOV-2002; 2002US-0428094P.
XX
XX (AMHP ) WYETH.
XX
XX O'coole M, Mounts WM, Shojasee N;
XX
XX WPI; 2004-441163/41.
XX
XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing
XX or treating autoimmune disorder, e.g. lupus or lupus nephritis.
XX
XX Example 11, Page 69, 83pp; English.
XX

```

XX The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.

XX SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSFC 14  
 Db 1 MIRSKITSVLSFC 14

RESULT 13

ADO71778 ADO71778 standard; protein; 1753 AA.

XX ADO71778;

XX 26-AUG-2004 (first entry)

XX BFLP0169 variant C1449Y, SEQ ID 17.

XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutuin.

XX Homo sapiens.  
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 1449 /note= "Wild-type C replaced with Y"

XX MO2004048521-A2.

XX 10-JUN-2004.

XX 21-NOV-2003; 2003WO-US037339.

XX 21-NOV-2002; 2002US-0428094P.

XX (AMHP ) WYETH.

XX O'coole M, Mounts WM, Shojaee N;

XX WPI; 2004-441163/41.

XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.

XX Example 16; Page 72; 83pp; English.

XX The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.

XX SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSFC 14  
 Db 1 MIRSKITSVLSFC 14

RESULT 14

ADO71769 ADO71769 standard; protein; 1753 AA.

XX ADO71769;

XX 26-AUG-2004 (first entry)

XX BFLP0169 variant L553I, SEQ ID 8.

XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutuin.

XX Homo sapiens.  
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 553 /note= "Wild-type L replaced with I"

XX MO2004048521-A2.

XX 10-JUN-2004.

XX 21-NOV-2003; 2003WO-US037339.

XX 21-NOV-2002; 2002US-0428094P.

XX (AMHP ) WYETH.

XX O'coole M, Mounts WM, Shojaee N;

XX WPI; 2004-441163/41.

XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.

XX Example 7; Page 66-67; 83pp; English.

XX The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.

XX SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSFC 14  
 Db 1 MIRSKITSVLSFC 14

RESULT 15

ADO71770 ADO71770 standard; protein; 1753 AA.

XX ADO71770;

XX 26-AUG-2004 (first entry)

XX BFLP0169 variant W634F, SEQ ID 9.

KW Anticardiolipin; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutain.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FT Misc-difference 634 /note= "Wild-type W replaced with F"  
 PT  
 XX  
 PN WO2004048521-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 21-NOV-2003; 2003WO-US037339.  
 XX  
 PR 21-NOV-2002; 2002US-0428094P.  
 XX  
 PA (AMHP ) WYETH.  
 XX  
 PI O'toole M, Mounts WM, Shojaae N,  
 XX  
 DR WPI, 2004-441163/41.  
 XX  
 PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 XX  
 PS Example 8; Page 67; 83pp; English.  
 XX  
 CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 for diagnosing and treating autoimmune disorder, preferably lupus or  
 lupus nephritis. They are also useful in the generation of antibodies  
 that bind immunospecifically to the nucleic acids. The present sequence  
 is a BFLP0169 variant.  
 CC  
 CC  
 SO Sequence 1753 AA;  
 Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MIRSKITSVLSFC 14  
 DB 1 MIRSKITSVLSFC 14  
 RESULT 16  
 ADO71774  
 ID ADO71774 standard; protein; 1753 AA.  
 XX  
 AC ADO71774;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE BFLP0169 variant K1084R, SEQ ID 13.  
 XX  
 KW Anticardiolipin; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutain.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FT Misc-difference 1084 /note= "Wild-type K replaced with R"  
 PT  
 XX  
 PN WO2004048521-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 21-NOV-2003; 2003WO-US037339.  
 XX

XX  
 PR 21-NOV-2002; 2002US-0428094P.  
 XX  
 PA (AMHP ) WYETH.  
 XX  
 PI O'toole M, Mounts WM, Shojaae N,  
 XX  
 DR WPI, 2004-441163/41.  
 XX  
 PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 XX  
 PS Example 12; Page 69-70; 83pp; English.  
 XX  
 CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 for diagnosing and treating autoimmune disorder, preferably lupus or  
 lupus nephritis. They are also useful in the generation of antibodies  
 that bind immunospecifically to the nucleic acids. The present sequence  
 is a BFLP0169 variant.  
 CC  
 CC  
 SO Sequence 1753 AA;  
 Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MIRSKITSVLSFC 14  
 DB 1 MIRSKITSVLSFC 14  
 RESULT 17  
 ADO71779  
 ID ADO71779 standard; protein; 1753 AA.  
 XX  
 AC ADO71779;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE BFLP0169 variant D1542Q, SEQ ID 18.  
 XX  
 KW Anticardiolipin; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutain.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FT Misc-difference 1542 /note= "Wild-type D replaced with Q"  
 PT  
 XX  
 PN WO2004048521-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 21-NOV-2003; 2003WO-US037339.  
 XX  
 PR 21-NOV-2002; 2002US-0428094P.  
 XX  
 PA (AMHP ) WYETH.  
 XX  
 PI O'toole M, Mounts WM, Shojaae N,  
 XX  
 DR WPI, 2004-441163/41.  
 XX  
 PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 XX  
 PS Example 17; Page 72-73; 83pp; English.  
 XX  
 CC The present invention relates to human BFLP0169 (ADO71763) and its coding

CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.

XX  
 SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

RESULT 18

ADO71763  
 ID ADO71763 standard; protein; 1753 AA.

XX ADO71763;

XX 26-AUG-2004 (first entry)

XX Human BFLP0169, SEQ ID 2.

XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;

KM kidney; lupus nephritis; autoimmune disorder; lupus.

XX Homo sapiens.

XX WO2004048521-A2.

XX 10-JUN-2004.

XX 21-NOV-2003; 2003WO-US037339.

XX 21-NOV-2002; 2002US-0428094P.

XX (AMHP) WYETH.

XX O'Loole M, Mounts WM, Shojaae N;

XX WPI; 2004-44163/41.

DR N-PSDB; ADO71762.

XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.

XX Claim 9; Page 7-8; 83pp; English.

CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids.

XX Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

RESULT 19  
 ADO71776

ID ADO71776 standard; protein; 1753 AA.

XX ADO71776;

XX 26-AUG-2004 (first entry)

XX BFLP0169 variant 11247V, SEQ ID 15.

XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutant.

XX Homo sapiens.

OS Synthetic.

XX Key location/Qualifiers

FT Misc-difference 1247 /note= "Wild-type I replaced with V"

XX WO2004048521-A2.

XX 10-JUN-2004.

XX 21-NOV-2003; 2003WO-US037339.

XX 21-NOV-2002; 2002US-0428094P.

XX (AMHP) WYETH.

XX O'Loole M, Mounts WM, Shojaae N;

XX WPI; 2004-44163/41.

XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.

XX Example 14; Page 70-71; 83pp; English.

CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.

XX Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

RESULT 20

ADH85577  
 ID ADH85577 standard; protein; 162 AA.

XX ADH85577;

XX 22-APR-2004 (first entry)

XX Enterococcus faecalis polypeptide #57.

XX Enterococcus faecalis infection; transcription regulatory element;

XX antibacterial.

XX Enterococcus faecalis.

XX US6617156-B1.

PD 09-SEP-2003.  
 XX 13-AUG-1998; 98US-00134000.  
 XX 15-AUG-1997; 97US-0055778P.  
 XX (DOHC/) DOUCETTE-STAMM L. A.  
 PA (BUSH/) BUSH D.  
 XX Doucette-Stamm LA, Bush D;  
 PI WPI: 2003-895394/82.  
 DR N-PSDB; ADH82172.  
 XX New nucleic acid comprising a sequence encoding an Enterococcus faecalis  
 PT polypeptide, useful for preparing a composition for diagnosing or  
 PR treating E. faecalis infection.  
 XX Disclosure; SEQ ID NO 3462; 193pp; English.  
 XX The invention relates to Enterococcus faecalis polynucleotides and  
 CC polypeptides. The invention also relates to a recombinant expression  
 CC vector comprising a polynucleotide operably linked to a transcription  
 CC regulatory element, a cell comprising a recombinant vector, a method for  
 CC producing an E. faecalis polypeptide, an isolated nucleic acid comprising  
 CC a sequence not given in the specification, a recombinant vector  
 CC comprising the nucleic acid and a cell comprising the recombinant vector.  
 CC The polynucleotides can be used to detect the presence of E. faecalis in  
 CC a sample. The sequences are useful for preparing a composition for  
 CC diagnosing or treating Enterococcus faecalis infection. This sequence  
 CC represents an E. faecalis polypeptide of the invention.  
 XX Sequence 162 AA;  
 SQ  
 QY Query Match 50.0%; Score 7; DB 7; Length 162;  
 DB Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 KITSVLS 12  
 DB 129 KITSVLS 135  
 RESULT 21  
 AAB79589  
 ID AAB79589 standard; protein; 285 AA.  
 XX AAB79589;  
 AC 30-APR-2001 (first entry)  
 DT Corynebacterium glutamicum SMP protein sequence SEQ ID NO:694.  
 DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:694.  
 XX Corynebacterium glutamicum; carbon metabolism and energy production;  
 KM SMP protein; sugar metabolism and oxidative phosphorylation protein;  
 KM fine chemical production; organic acid; pyrimidine base; nucleoside;  
 KM nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KM nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KM carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
 KM diagnosis; Corynebacterium diptheriae; evolutionary study.  
 XX Corynebacterium glutamicum.  
 OS WO200100844-A2.  
 XX 04-JAN-2001.  
 PD 23-JUN-2000; 2000MO-IB000943.  
 XX 25-JUN-1999; 99US-0141031P.  
 PR 08-JUN-1999; 99DE-01031412.  
 PR 08-JUN-1999; 99DE-01031413.  
 PR 08-JUN-1999; 99DE-01031419.  
 PR 08-JUN-1999; 99DE-01031419.

PR 08-JUN-1999; 99DE-01031420.  
 PR 08-JUN-1999; 99DE-01031424.  
 PR 08-JUN-1999; 99DE-01031428.  
 PR 08-JUN-1999; 99DE-01031431.  
 PR 08-JUN-1999; 99DE-01031433.  
 PR 08-JUN-1999; 99DE-01031434.  
 PR 08-JUN-1999; 99DE-01031435.  
 PR 08-JUN-1999; 99DE-01031510.  
 PR 08-JUN-1999; 99DE-01031562.  
 PR 08-JUN-1999; 99DE-01031634.  
 PR 09-JUN-1999; 99DE-01032180.  
 PR 09-JUN-1999; 99DE-01032227.  
 PR 09-JUN-1999; 99DE-01032230.  
 PR 09-JUN-1999; 99US-0143208P.  
 PR 14-JUN-1999; 99DE-01032924.  
 PR 14-JUN-1999; 99DE-01032973.  
 PR 14-JUN-1999; 99DE-01033005.  
 PR 27-AUG-1999; 99DE-01040765.  
 PR 31-AUG-1999; 99US-0151572P.  
 PR 03-SEP-1999; 99DE-01042076.  
 PR 03-SEP-1999; 99DE-01042079.  
 PR 03-SEP-1999; 99DE-01042086.  
 PR 03-SEP-1999; 99DE-01042087.  
 PR 03-SEP-1999; 99DE-01042088.  
 PR 03-SEP-1999; 99DE-01042095.  
 PR 03-SEP-1999; 99DE-01042123.  
 PR 03-SEP-1999; 99DE-01042125.  
 XX (BADI ) BASF AG.  
 PA Pompejus M, Kroege B, Schroeder H, Zelder O, Habehauer G;  
 PI WPI: 2001-061975/07.  
 DR N-PSDB; AAF71706.  
 XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar  
 PT metabolism and oxidative phosphorylation protein for production or  
 PR modulation of production of fine chemicals e.g. amino acids,  
 PT carbohydrates or enzymes.  
 XX Claim 20; Page 1129-1130; 1246pp; English.  
 PS AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar  
 CC metabolism and oxidative phosphorylation (SMP) proteins given in AAB79243  
 CC to AAB 79633 which are involved in carbon metabolism and energy  
 CC production. The C. glutamicum SMP gene can be used in vectors (II) for  
 CC expression in host cells and production or modulation of production of  
 CC fine chemicals, such as, an organic acid, a proteinogenic or  
 CC nonproteinogenic amino acid (preferred), a purine or pyrimidine base, a  
 CC nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty acid,  
 CC a diol, a carbohydrate, an aromatic compound, a vitamin, a cofactor, a  
 CC polyketide, or an enzyme. The presence of (I) or SMP proteins (III)  
 CC encoded by them are used for diagnosing the presence or activity of  
 CC Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells  
 CC containing them are used to map genomes of organisms related to C.  
 CC glutamicum, identify and localise C. glutamicum sequences of interest, in  
 CC evolutionary studies, in determining SMP protein regions required for  
 CC function, in modulating SMP protein activity, in modulating the  
 CC metabolism of sugars, and in modulating high-energy molecule production  
 CC in a cell (I.e. ATP, NADPH)  
 CC XX  
 SQ Sequence 285 AA;  
 QY Query Match 50.0%; Score 7; DB 4; Length 285;  
 DB Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 ITSVLSP 13  
 DB 139 ITSVLSP 145  
 RESULT 22  
 AAB67357

ID ABM67357 standard; protein; 366 AA.  
 XX ABM67357;  
 AC  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Photorhabdus luminescens protein sequence #454.  
 XX  
 KM Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KM detection; food; gene expression; plant; animal; microorganism; toxin;  
 KM antibiotic; biopesticide; virulence factor; disease model; plague;  
 KM whooping cough.  
 OS  
 PN Photorhabdus luminescens.  
 XX WO200294867-A2.  
 XX  
 PD 28-NOV-2002.  
 XX  
 PF 07-FEB-2002; 2002WO-1B003040.  
 XX  
 PR 07-FEB-2001; 2001FR-00001659.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A,  
 PI Buchrieser C;  
 DR WPI; 2003-146459/14.  
 XX  
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX  
 PS Claim 2; SEQ ID NO 454; 1205bp; French.  
 XX  
 CC The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins  
 XX  
 SO Sequence 366 AA:  
 Query Match 50.0%; Score 7; DB 6; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 RSKITS 9  
 DB 354 RSKITS 360  
 RESULT 23  
 ADO25476  
 ID ADO25476 standard; protein; 430 AA.  
 XX

AC ADO25476;  
 XX  
 DT 12-AUG-2004 (first entry)  
 XX  
 DE E.faecalis adenylosuccinate lyase purB protein Seg151.  
 XX  
 KM antimicrobial; microbial disease; drug composition; vaccine;  
 KM bacterial infection; antibacterial; food preservative.  
 XX  
 OS Enterococcus faecalis.  
 XX  
 PN WO2004041854-A2.  
 XX  
 PD 21-MAY-2004.  
 XX  
 PF 05-NOV-2003; 2003WO-CA001671.  
 XX  
 PR 05-NOV-2002; 2002US-0423757P.  
 PR 05-NOV-2002; 2002US-0423758P.  
 PR 05-NOV-2002; 2002US-0423791P.  
 PR 05-NOV-2002; 2002US-0423832P.  
 PR 05-NOV-2002; 2002US-0423875P.  
 PR 05-NOV-2002; 2002US-0423915P.  
 PR 06-NOV-2002; 2002US-0424362P.  
 PR 06-NOV-2002; 2002US-0424367P.  
 PR 06-NOV-2002; 2002US-0424370P.  
 PR 06-NOV-2002; 2002US-0424373P.  
 PR 06-NOV-2002; 2002US-0424376P.  
 PR 06-NOV-2002; 2002US-0424389P.  
 PR 07-NOV-2002; 2002US-0424502P.  
 PR 07-NOV-2002; 2002US-0424651P.  
 PR 07-NOV-2002; 2002US-0424664P.  
 PR 07-NOV-2002; 2002US-0424665P.  
 PR 08-NOV-2002; 2002US-0424968P.  
 PR 08-NOV-2002; 2002US-0425085P.  
 PR 08-NOV-2002; 2002US-0425118P.  
 PR 08-NOV-2002; 2002US-0425125P.  
 PR 08-NOV-2002; 2002US-0425162P.  
 PR 08-NOV-2002; 2002US-0425201P.  
 PR 12-MAR-2003; 2003US-0453914P.  
 PR 12-MAR-2003; 2003US-0454021P.  
 PR 12-MAR-2003; 2003US-0454128P.  
 PR 12-MAR-2003; 2003US-0454193P.  
 PR 13-MAR-2003; 2003US-0454215P.  
 PR 13-MAR-2003; 2003US-0454218P.  
 PR 13-MAR-2003; 2003US-0454487P.  
 PR 13-MAR-2003; 2003US-0454507P.  
 PR 13-MAR-2003; 2003US-0454536P.  
 PR 14-MAR-2003; 2003US-0455010P.  
 PR 14-MAR-2003; 2003US-0455036P.  
 PR 14-MAR-2003; 2003US-0455054P.  
 PR 14-MAR-2003; 2003US-0455082P.  
 PR 17-MAR-2003; 2003US-0455191P.  
 PR 17-MAR-2003; 2003US-0455192P.  
 PR 17-MAR-2003; 2003US-0455334P.  
 PR 17-MAR-2003; 2003US-0455335P.  
 PR 17-MAR-2003; 2003US-0455343P.  
 XX  
 PA (AFPI-) AFFINITUM PHARM INC.  
 XX  
 PI Edwards A, Dharmas A, Vedadi M, Domagala M, Nethery K,  
 PI Manouky K, Pinder B, Alam MZ, Ng I, Virag C, Houston S,  
 PI McDonald M, Buzadzija K;  
 XX  
 DR WPI; 2004-400642/37.  
 DR N-PSDB; ADO25475.  
 XX  
 PT Bacterial polypeptide composition useful for treating bacterial  
 PT infection, has isolated, recombinant bacterial polypeptide such as GTP-  
 PT binding protein Era from Pseudomonas aeruginosa or adenylosuccinate lyase  
 PT from Enterococcus faecalis.  
 XX

PS Claim 40; SEQ ID NO 151; 566bp; English.

XX This invention relates to the identification of novel protein targets for  
CC the development of antimicrobial drugs against pathogenic bacteria.  
CC Specifically, it refers to recombinant proteins derived from  
CC *Staphylococcus aureus*, *Helicobacter pylori*, *Streptococcus pneumoniae*,  
CC *Escherichia coli*, *Enterococcus faecalis* or *Pseudomonas aeruginosa*. The  
CC present invention describes providing a three-dimensional structure for  
CC these crystallised proteins to identify a potential modulator for the  
CC prevention or treatment of microbial diseases. Furthermore, contacting a  
CC protein with a modulator can be useful for assaying protein activity and  
CC hence its viability in drug composition or vaccine. Accordingly, such  
CC compositions can be useful for treating bacterial infections, developing  
CC antibacterial agents useful as food preservatives or treating food  
CC products to eliminate potential pathogens. This polypeptide sequence is a  
CC bacterial protein target of the invention.

SO Sequence 430 AA;

Query Match 50.0%; Score 7; DB 8; Length 430;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 KITSVLS 12  
Db 397 KITSVLS 403

RESULT 24

AAU35142  
ID AUA35142 standard; protein; 431 AA.

XX AAU35142;

DT 13-FEB-2002 (first entry)

DE Enterococcus faecalis cellular proliferation protein #429.

XX Antisense; prokaryotic cellular proliferation protein; antibiotic;

KW antibacterial; drug design.

XX Enterococcus faecalis.

OS WO200170955-A2.

XX 27-SEP-2001.

PD 21-MAR-2001; 2001WO-US009180.

PR 21-MAR-2001; 2000US-0191078P.

PR 23-MAY-2000; 2000US-0206848P.

PR 26-MAY-2000; 2000US-0207727P.

PR 23-OCT-2000; 2000US-0242578P.

PR 27-NOV-2000; 2000US-0253625P.

PR 22-DEC-2000; 2000US-0257931P.

PR 16-FEB-2001; 2001US-0269308P.

XX (ELIT-) ELITRA PHARM INC.

PI Haasebeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR N-PSDB; AAS53001.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids.

XX Example 3; SEQ ID NO 10735; 511bp; English.

CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*

CC coli, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,  
CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also  
CC useful for the identification of potential new targets for antibiotic  
CC development. The antisense nucleic acids can also be used to identify  
CC proteins used in proliferation, to express these proteins, and to obtain  
CC antibodies capable of binding to the expressed proteins. The proteins can  
CC be used to screen compounds in rational drug discovery programmes. The  
CC antisense nucleic acid sequence is also useful to screen for homologous  
CC nucleic acids which are required for cell proliferation in a wide variety  
CC of organisms. The present sequence represents an essential prokaryotic  
CC cellular proliferation protein. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

SO Sequence 431 AA;

Query Match 50.0%; Score 7; DB 4; Length 431;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 KITSVLS 12  
Db 398 KITSVLS 404

RESULT 25

ABU29114  
ID ABU29114 standard; protein; 431 AA.

XX ABU29114;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #14641.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Enterococcus faecalis.

OS WO200277183-A2.

XX 03-OCT-2002.

PD 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haasebeck R, Ohlsen KL, Zyskind JW,

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;

XX WPI; 2003-028926/02.

DR N-PSDB; AKA32984.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

XX isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 57038; 1766bp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated





AC AAG92745;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE C glutamicum protein fragment SEQ ID NO: 6499.  
 XX  
 KW Corynebacterium; amino acid synthetase; vitamin; saccharide;  
 XX organic acid synthetase.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EPI108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-00127688.  
 XX  
 PR 16-DEC-1999; 99DP-00377484.  
 XX  
 PR 07-APR-2000; 2000JP-00159162.  
 XX  
 PR 03-AUG-2000; 2000JP-00280988.  
 XX  
 PA (KYOM ) KYOMA HAKKO KOGYO KK.  
 XX  
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H,  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 DR WPI; 2001-376931/40.  
 DR N-PSDB; AAB67964.  
 XX  
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analyzing  
 PT expression profile or pattern of a gene and identifying homologous gene.  
 XX  
 PS Claim 17: SEQ ID NO 6499; 246bp + Sequence listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and analysing  
 CC the expression profile or expression pattern of a gene derived from  
 CC Coryneform bacterium, and identifying a homologue of a gene derived from  
 CC coryneform bacterium. Coryneform bacteria are useful for producing amino  
 CC acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described in the  
 CC exemplification of the invention. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from the European Patent Office  
 XX  
 SQ Sequence 1019 AA;  
 SQ  
 Query Match 50.0%; Score 7; DB 4; Length 1019;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 ITSVLSP 13  
 DB 139 ITSVLSP 145  
 DB  
 RESULT 28  
 ABB79444  
 ID ABB79444 standard; protein; 1019 AA.  
 XX  
 AC ABB79444;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 DE Corynebacterium Nuov SEQ ID NO 12.  
 XX  
 KW Corynebacterium; ctad; cytochrome a3 oxidase subunit I; ctAB;  
 KW cytochrome a3 subunit III; qcrC; cytochrome c1; qcrA; nuov; nuov;  
 KW Rieske iron-sulphur protein; qcrB; cytochrome b; metallothionein;  
 KW organic acid; mono-valent alcohol; poly-valent alcohol; ATP synthetase;

KW NADH(reduced nicotinamide-adenine dinucleotide)-dehydrogenase subunit;  
 KW electron-transport chain; respiration; enzyme.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1  
 FT /note= "Encoded by GTC"  
 XX  
 PN WO200222799-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 08-SEP-2001; 2001WO-DE003485.  
 XX  
 PR 14-SEP-2000; 2000US-0232514P.  
 XX  
 PA (KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.  
 PA (DEGS ) DEGUSA AG.  
 XX  
 PI Bott M, Nieblisch A, Bathe B, Marx A, Hermann T;  
 DR WPI; 2002-393962/42.  
 DR N-PSDB; ABL58732.  
 XX  
 PT New genes from coryneform bacteria useful, when modulated, for  
 PT controlling production of metabolites, e.g. amino acids or vitamins.  
 XX  
 PS Claim 6; Page 82-85; 98pp; German.  
 XX  
 CC The invention relates to polynucleotides (ABL58727-ABL58734) that  
 CC represent the ctad (cytochrome a3 oxidase subunit I) ctAB (cytochrome  
 CC a3 subunit III), qcrC (cytochrome c1), qcrA (Rieske iron-sulphur  
 CC protein), qcrB (cytochrome b) and nuov, V and W (NADH (reduced  
 CC nicotinamide-adenine dinucleotide)-dehydrogenase subunits U, V and W)  
 CC genes of coryneform bacteria and the encoded proteins (ABB79439-  
 CC ABB79446). Bacteria in which activities of one or more of the new genes  
 CC have been reduced or increased are useful for producing metabolites,  
 CC particularly amino acids, organic acids, nucleotides, nucleosides and  
 CC mono- or poly-valent alcohols. The new genes can also be used as  
 CC hybridisation probes to identify related nucleic acids. Altering the  
 CC activity of the new gene products makes it possible to modulate ATP  
 CC (adenosine triphosphate) synthesis by electron-transport phosphorylation  
 CC via the respiratory chain and thus to control metabolite synthesis  
 XX  
 SQ Sequence 1019 AA;  
 SQ  
 Query Match 50.0%; Score 7; DB 5; Length 1019;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 ITSVLSP 13  
 DB 139 ITSVLSP 145  
 DB  
 RESULT 29  
 ADL65625  
 ID ADL65625 standard; protein; 1019 AA.  
 XX  
 AC ADL65625;  
 XX  
 DT 20-MAY-2004 (first entry)  
 XX  
 DE C. glutamicum membrane transport and synthesis-associated protein #99.  
 XX  
 KW fine chemical production; lysine production; nucleotide; nucleoside;  
 KW lipid; fatty acid; diol; carbohydrate; aromatic compound; vitamin;  
 KW co-factor; enzyme; food; animal feed; cosmetic; pharmaceutical; gene;  
 KW membrane synthesis; membrane transport.  
 XX  
 OS Corynebacterium glutamicum.

PN DE10154179-A1.  
 XX 08-MAY-2003.  
 PD 05-NOV-2001; 2001DE-01054179.  
 XX 05-NOV-2001; 2001DE-01054179.  
 XX (BADI ) BASF AG.  
 XX  
 PI Zelder O, Pompejus M, Schroeder H, Kroegeer B, Kloppeggge C,  
 PI Haberhauser G;  
 XX WPI; 2003-442553/42.  
 DR N-P8DB; ADL65624.  
 XX  
 PT New nucleic acid encoding variant forms of membrane construction and  
 PT transport proteins, useful for production of fine chemicals from  
 PT microorganisms, e.g. nucleotides, lipids and especially lysine.  
 XX  
 PS Claim 1; Page; 20pp; German.  
 XX  
 CC This invention describes novel polynucleotides that encode membrane  
 CC synthase and membrane transport proteins from *Corynebacterium glutamicum*.  
 CC The polynucleotides are isolated from a nucleic acid library of C.  
 CC glutamicum then mutated at the specified positions, cloned and expressed  
 CC by standard methods. Cells, especially *Corynebacterium glutamicum*,  
 CC containing vectors that express the polynucleotides are used for  
 CC production of fine chemicals, preferably amino acids and specifically  
 CC lysine, but more generally nucleotides, nucleosides, lipids, fatty acids,  
 CC diols, carbohydrates, aromatic compounds, vitamins, co-factors and  
 CC enzymes. These are useful in the food, animal feed, cosmetics and  
 CC pharmaceutical industries. The polynucleotides, optionally as primers and  
 CC probes, can also be used for identification and classification of C.  
 CC glutamicum and related species, e.g. for diagnosis, for genomic mapping,  
 CC functional or evolutionary studies, gene manipulation and modulation of  
 CC metabolic activity. Cells that containing the polynucleotides of the  
 CC invention may produce fine chemicals in better yields, with higher  
 CC productivity and/or more efficiently. NOTE: This sequence is not  
 CC represented in the printed specification but is available in electronic  
 CC format. The sequence represented in this record has been obtained from  
 CC WO2003040292.  
 XX  
 SQ Sequence 1019 AA;  
 Query Match 50.0%; Score 7; DB 7; Length 1019;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 31-AUG-2000; 2000MO-US024034.  
 PF 01-SEP-1999; 99US-0151980P.  
 PR 13-MAR-2000; 2000US-0188944P.  
 XX  
 PA (UYP1-) UNIV PITTSBURGH.  
 XX  
 PI Robbins PD, M1 Z, Fritzell R, Glorioso JC, Gambotto A;  
 PI WPI; 2001-273309/28.  
 XX  
 PT Peptides that facilitate uptake and cytoplasmic and/or nuclear transport  
 PT of proteins, DNA and viruses, useful, e.g. for facilitating uptake of  
 PT antigens in immunogenic compositions.  
 XX  
 PS Claim 1; Page 115; 129pp; English.  
 XX  
 CC The present invention provides the sequences of 75 peptides which  
 CC facilitate the uptake and transport of viruses, proteins and nucleic  
 CC acids. These internalising peptides can be used for transport into the  
 CC cytoplasm or the nucleus. They are useful for facilitating uptake into  
 CC the cell, inducing apoptosis, for example in the treatment of arthritis  
 CC and cancer, to expand a population of stem cells or differentiated cells,  
 CC to stimulate cell differentiation, facilitate the integration of AAV into  
 CC the genome of a cell, and to stimulate an immune response, for example in  
 CC the case of a HIV vaccine. The present sequence is one of the peptides of  
 CC the invention  
 XX  
 SQ Sequence 12 AA;  
 Query Match 42.9%; Score 6; DB 4; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IRKSKI 7  
 DB 3 IRKSKI 8  
 ID ADA86830  
 ID ADA86830 standard; peptide; 12 AA.  
 AC ADA86830;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Internalised peptide pep10 SEQ ID NO:10.  
 XX  
 KW Internalising peptide; cytosolic; antiinflammatory; immunomodulator;  
 KW antiarthritic; cytoplasmic transport; nuclear transport;  
 KW peptide-cargo complex; apoptosis; arthritic; tumour; differentiation;  
 KW immune response; vaccine; inflammation; necrosis; transplantation;  
 KW cystic fibrosis; lung inflammation; gene therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2003068942-A2.  
 XX  
 PD 21-AUG-2003.  
 XX  
 PF 12-FEB-2003; 2003WO-US004632.  
 XX  
 PR 13-FEB-2002; 2002US-00075869.  
 XX  
 PA (UYP1-) UNIV PITTSBURGH.  
 PI Robbins PD, M1 Z, Fritzell R, Glorioso JC, Gambotto A, Mai JC;  
 PI WPI; 2003-697526/66.  
 DR  
 XX  
 PT New internalising peptides, useful for facilitating the delivery, uptake

PT and cytoplasmic and/or nuclear transport of proteins, DNA or viruses into  
PT a target cell, for inducing apoptosis in arthritic or tumor cells, or in  
PT gene therapy.

PS Example 3, Page 18, 171pp, English.

CC The present invention describes an internalising peptide (I) comprising  
CC any one of 14 fully defined amino acid sequences (designated PI-P14, see  
CC ADA8896 to ADA8906, and ADA8917 to ADA8919). (I) has cytosolic,  
CC anti-inflammatory, immunomodulator and antirheumatic activities. The  
CC internalising peptides are useful for facilitating the delivery, uptake  
CC and cytoplasmic and/or nuclear transport of cargo, e.g. proteins, DNA or  
CC viruses, into a target cell. The internalising peptides and peptide-cargo  
CC complexes in cells (e.g. arthritic cells or tumour cells), expanding a  
CC population of stem cell or differentiated cells, stimulating the  
CC differentiation of a population of stem cells, facilitating the  
CC integration of adeno-associated virus DNA into the genome of a cell,  
CC stimulating or eliciting an immune response in a subject, facilitating  
CC the delivery of immunogens (e.g. vaccines), inhibiting the inflammatory  
CC process, protecting tissue from apoptosis or necrosis during tissue  
CC isolation prior to transplantation, facilitating transfer of proteins and  
CC peptides to the lung for the treatment of cystic fibrosis or lung  
CC inflammation, or in gene therapy. The present sequence represents a  
CC peptide used in the exemplification of the present invention.

CC Sequence 12 AA:

Query Match 42.9%; Score 6; DB 7; Length 12;  
Best Local Similarity 100.0%; Pred. No. 12;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IRRSKI 7  
| | | | |  
DB 3 IRRSKI 8

RESULT 32

ABG78459  
ID ABG78459 standard; protein; 17 AA.

AC ABG78459;

DT 15-NOV-2002 (first entry)

DE HLRSII, transmembrane domain #3.

XX Human; human leucine-rich repeat small intestine I; HLRSII, asthma;  
KW proliferative disorder; gastrointestinal disorder; renal disorder;  
KW neural disorder; reproductive disorder; calcium regulation; apoptosis;  
KW immune system; anaemia; human immune deficiency virus; HIV; cancer;  
KW blood coagulation disorder; autoimmune disorder; allergic reaction;  
KW inflammatory condition; cardiovascular disorder; ischaemia;  
KW neurological disorder; infectious disease; cytokine production;  
KW expressed sequence tag; EST.

XX Homo sapiens.

OS WO200261086-A2.

PN 08-AUG-2002.

PD 20-DEC-2001; 2001WO-US049739.

PP 22-DEC-2000; 2000US-0257774P.

PR (BRIM ) BRISTOL-MYERS SQUIBB CO.

PA Feder J, Ramanathan C, Muntler G;

PI WPI, 2002-619252/66.

XX New isolated nucleic acid molecules encoding HLRSII polypeptides, or

PT their fragments and homologues, useful for preventing, treating and  
PT ameliorating medical conditions, e.g. proliferative, gastrointestinal, or  
PT renal disorders.

PS Disclosure, Page 225, 336pp, English.

CC The invention relates to isolated nucleic acid molecules (I) encoding  
CC human leucine-rich repeat small intestine I (HLRSII) polypeptides. The  
CC nucleic acid molecules and polypeptides are useful for preventing,  
CC treating and ameliorating medical conditions, such as proliferative,  
CC gastrointestinal, renal, neural, or reproductive disorders, or disorders  
CC related to aberrant calcium regulation or apoptosis modulation, either  
CC directly or indirectly. They are also useful for treating, preventing  
CC and/or diagnosing diseases, disorders and/or conditions of: immune system  
CC by activating or inhibiting the proliferation, differentiation, or  
CC mobilisation of immune cells; haematopoietic cells e.g. thrombocytopenia,  
CC anaemia; immunologic deficiency syndromes, e.g. human immune deficiency  
CC virus (HIV) infection, HTLV-BLV infection; blood coagulation disorders,  
CC e.g. arterial thrombosis; autoimmune disorders, e.g. Addison's disease,  
CC myasthenia gravis; asthma or allergic reactions; inflammatory conditions,  
CC e.g. chronic prostatitis, sepsis; proliferative disorders, e.g. cancer;  
CC cardiovascular disorders, e.g. arrhythmia, myocardial ischaemia;  
CC aneurysms; neurological disorders, e.g. Alzheimer's disease, Huntington's  
CC chorea; infectious diseases, e.g. measles, mumps, pneumonia, or viral,  
CC bacterial, and fungal infections. The HLRSII polypeptides are useful for  
CC modulating cytokine production, antigen presentation, or other processes  
CC such as boosting immune responses. ABG78454-ABG78474 represent HLRSII  
CC amino acid sequences and related amino acid sequences of the invention

CC Sequence 17 AA:

Query Match 42.9%; Score 6; DB 5; Length 17;  
Best Local Similarity 100.0%; Pred. No. 17;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
| | | | |  
DB 7 ITSVLS 12

RESULT 33

AAG00649  
ID AAG00649 standard; protein; 77 AA.

AC AAG00649;

DT 06-OCT-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 4730.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping.

XX Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PP 21-FEB-2000; 2000EP-00200610.

PR 26-FEB-1999; 99US-0122487P.

PA (GEST ) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

PA WPI, 2000-500381/45.

DR N-PSDB, AAC00655.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX  
PS Claim 13; SEQ ID NO 4730; 71bp + Sequence Listing; English.

CC The present sequence is a polypeptide encoded by one of a large number of  
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were  
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different  
CC tissues. EST sequences usually correspond mainly to the 3' untranslated  
CC region (UTR) of the mRNA because they are often obtained from oligo-dT  
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA  
CC sequences derived from the 5' ends of mRNAs and even in those cases where  
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely  
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can  
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs  
CC are also used in diagnostic, forensic, gene therapy and chromosome  
CC mapping procedures. They are used to obtain upstream regulatory sequences  
CC and to design expression and secretion vectors

XX  
SQ Sequence 77 AA:

Query Match 42.9%; Score 6; DB 3; Length 77;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KSKITS 9  
|||||  
67 KSKITS 72

DB 67 KSKITS 72

RESULT 34  
ABG06877  
ID ABG06877 standard; protein; 85 AA.

XX  
AC ABG06877;  
XX  
DT 13-FEB-2002 (first entry)

XX  
DE Novel human diagnostic protein #6868.

XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX  
OS Homo sapiens.

XX  
PN WO200175067-A2.

XX  
PD 11-OCT-2001.

XX  
PF 30-MAR-2001; 2001WO-US008631.

XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.

XX  
PA (HYSB-) HYSEQ INC.

XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS71064.

XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostic, forensic, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX  
PS Claim 20; SEQ ID NO 37236; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (II) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is

CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG3037 represent novel human diagnostic  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 85 AA:

Query Match 42.9%; Score 6; DB 4; Length 85;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRKSK 6  
|||||  
49 MIRKSK 54

DB 49 MIRKSK 54

RESULT 35  
AAU39937  
ID AAU39937 standard; protein; 90 AA.

XX  
AC AAU39937;  
XX  
DT 13-FEB-2002 (first entry)

XX  
DE Propionibacterium acnes immunogenic protein #833.

XX  
KW SAPRO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.

XX  
OS Propionibacterium acnes.

XX  
PN WO200181581-A2.

XX  
PD 01-NOV-2001.

XX  
PF 20-APR-2001; 2001WO-US012865.

XX  
PR 21-APR-2000; 2000US-0199047P.  
PR 02-JUN-2000; 2000US-0208841P.  
PR 07-JUL-2000; 2000US-0216747P.

XX  
PA (COR-) CORIXA CORP.

XX  
PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
DR WPI; 2001-616774/71.  
DR N-PSDB; AAS59510.

XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.

XX  
PS Example 1; SEQ ID NO 1132; 1069bp; English.

CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPRO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central

CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting the  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 90 AA;

Query Match 42.9%; Score 6; DB 4; Length 90;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ITSVLS 12  
DB 5 ITSVLS 10

RESULT 36  
ABM36456  
ID ABM36456 standard; protein; 90 AA.  
XX  
AC ABM36456;  
XX  
DT 20-OCT-2003 (first entry)  
XX  
DE Propionibacterium acnes predicted ORF-encoded polypeptide #1132.  
XX  
KM Acne vulgaris; anti-seborrheic; dermatological; antibacterial;  
XX immunostimulant; immune response; vaccine.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO2003033515-A1.  
XX  
PD 24-APR-2003.  
XX  
PF 11-OCT-2002; 2002MO-US032727.  
XX  
PR 15-OCT-2001; 2001US-00978825.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Mitcham JL, Skeiky YAM, Persing DH, Bhattacha A, Maisonneuve JL,  
XX Zhang Y, Wang S, Jen S, Lodes MT, Benson DR, Jones R, Carter D,  
XX Barth B, Valliave-Douglas J;  
XX  
DR MPI, 2003-381789/36.  
XX  
DR N-PSDB; ACP64439.  
XX  
XX

PT New Propionibacterium acnes polypeptides and polynucleotides encoding the  
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a P. acnes protein.  
XX  
PS Example 1; SEQ ID NO 1132; 1481bp; English.

CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC encoding a Propionibacterium acnes protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
CC immunogenic fragments of P. acnes polypeptides. The invention  
CC additionally encompasses expression vectors and host cells comprising a  
CC polynucleotide of the invention; antibodies against polypeptides of the  
CC invention; fusion proteins comprising a polypeptide of the invention; a  
CC method for stimulating an immune response specific for a P. acnes  
CC polypeptide and an isolated T cell population comprising T cells prepared

CC via this method, a vaccine composition (comprising P. acnes polypeptides,  
CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
CC antigen-presenting cells that express the polypeptide); a method and kit  
CC for detecting or determining the presence or absence of P. acnes in a  
CC patient; and a method for inhibiting the development of P. acnes in a  
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
CC proteins, T cell populations or antigen-presenting cells that express the  
CC polypeptides are useful for diagnosing, preventing or treating acne  
CC vulgaris, or for stimulating an immune response specific for a P. acnes  
CC protein. The polynucleotides can also be used as probes or primers for  
CC nucleic acid hybridisation. The vaccine composition is useful for the  
CC stimulation of an immune response against P. acnes, or for treating acne,  
CC and the kit is useful for performing a diagnostic assay. The present  
CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
CC reading frame) contained within the P. acnes polynucleotides of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 90 AA;

Query Match 42.9%; Score 6; DB 6; Length 90;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ITSVLS 12  
DB 5 ITSVLS 10

RESULT 37  
ABU17550  
ID ABU17550 standard; protein; 130 AA.  
XX  
AC ABU17550;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #3077.  
XX  
KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
XX  
OS Bacillus anthracis.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002MO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
XX  
PR 06-SEP-2001; 2001US-00948993.  
XX  
PR 25-OCT-2001; 2001US-0349223P.  
XX  
PR 08-FEB-2002; 2002US-00072851.  
XX  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX  
XX

PA (ELIT-) ELITRA PHARM INC.  
XX  
XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW,  
PI Wall D, Trawick JD, Carr GT, Yamamoto R, Foreyth RA, Xu HH,  
XX  
XX  
DR MPI, 2003-029926/02.  
XX  
DR N-PSDB; ACA21420.  
XX  
XX

PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX

PS Claim 25; SEQ ID NO 45474; 1766bp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 623 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:

(1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

**SQ Sequence 130 AA;**

Query Match	42.9%	Score 6	DB 6	Length 130
Similarity	100.0%	Pred. No. 97		
Best Local				
Matches 6	Conservative 0	Mismatches 0	Gaps 0	

QY	3	RKSKIT	8
Db	28	RKSKIT	33

```

RESULT 38
ADC94908
ID ADC94908 standard; protein; 133 AA

```

AC ADC94908,

DT 01-JAN-2004 (first entry)

DE E. faecium protein sequence SEQ ID 4535.

KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound, abdominal-pelvic infection.

**OS Enterococcus faecium.**

PN US6583275-B1

PD 24-JUN-2003

PF 30-JUN-1998, 98US-00107532.

PR 02-JUL-1997, 97US-0051571P.

XX

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Doucette-Stamm LA, Bush D;

DR WPI; 2003-799836/75.

DR N-PSDB; ADC91254.

PT New isolated nucleic acid derived from *Enterococcus faecium* encoding an  
PT *Enterococcus faecium* polypeptide useful for detection, prevention and

PT treatment of a pathological condition resulting from a bacterial infection.

PS Example 1; SEQ ID NO 4535; 243pp; English

The invention relates to an isolated nucleic acid derived from *Enterococcus faecium* encoding an *Enterococcus faecium* polypeptide having one of 10 fully defined sequences given in the (or comprising 40 sequential nucleotides chosen from any of the nucleic acids, its complement or sequences hybridising to it). Also included are a recombinant vector comprising the nucleic acid operably linked to transcription regulatory element, a cell comprising the vector and a single-stranded probe comprising the nucleic acid. The nucleic acids are chosen from 3654 disclosed sequences encoding 3654 disclosed proteins. The nucleic acids is useful for diagnosing pathological conditions resulting from *E. faecium* bacterial infection (e.g. urinary tract infection, bacteraemia, endocarditis, wounds and abdominal-pelvic infection) and for screening drugs such as agonists and antagonists. The nucleic acid is useful for recombinant production of Candida albicans - derived peptides or antisense polypeptides. Pharmaceutical compositions and vaccines containing the nucleic acid are useful for preventing or treating *Enterococcus faecium* infections. The present sequence represents one if the disclosed *E. faecium* proteins.

SQ Sequence 133 AA;

Query Match	42.9%	Score 6	DB 7	Length 133
Best Local Similarity	100.0%	Pred. No. 99		
Matches 6	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	4	KSKITS	9
Db	120	KSKITS	125

RESULT 39	
ADF04124	
ID	ADF04124 standard; protein; 135 AA

AC ADF04124

DT 12-FEB-2004 (first entry)

DE Bacterial polypeptide #237.

**KW Proteus mirabilis infection; bacterial infection; antibacterial, immunostimulant.**

## OS Proteus mirabilis

PN US6605709-B1

PD 12-AUG-2003

PF 05-APR-2000; 2000US-00543681.

PR 09-APR-1999; 99US-0128706P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Breton GL;

DR WPI; 2003-895291/82.

2000

PT New *Proteus mirabilis* polypeptides and polymucleotides, useful as reagents for diagnosis of bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, or as biocontrol agents for plants.

PS Disclosure; SEQ ID NO 4409; 870pp; English

CC The invention relates to new Proteus mirabilis polypeptides and



CC polynucleotides. The invention also relates to antibodies against the  
 CC polypeptides, methods for producing the polypeptides, a method of  
 CC generating vaccines for immunising an individual against *P. mirabilis*, a  
 CC method for evaluating a compound for the ability to bind a *P. mirabilis*  
 CC activity, The polypeptides and polynucleotides are useful as molecular  
 CC targets for diagnosing, preventing and treating pathological conditions  
 CC resulting from bacterial infection, as reagents for diagnosis of  
 CC bacterial diseases, as components of antibacterial vaccines, as targets  
 CC for antibacterial drugs or as bio-control agents for plants. This  
 CC sequence represents a *Proteus mirabilis* polypeptide of the invention.  
 SQ Sequence 135 AA;  
 Query Match 42.9%; Score 6; DB 7; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 ITSVL 11  
 DB 3 ITSVL 8  
 RESULT 40  
 ABP10237  
 ID ABP10237 standard; protein; 139 AA.  
 AC ABP10237;  
 XX  
 DT 24-JUN-2002 (first entry)  
 DE Human ORFX protein sequence SEQ ID NO:20456.  
 XX  
 KM Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KM hyperproliferative disorder; psoriasis; benign tumor; haemorrhage;  
 KM degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KM cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KM hypertension; hypothyroidism; cholesterol ester storage disease;  
 KM immune deficiency; immune disorder; infectious disease;  
 KM autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KM myasthenia gravis.  
 OS Homo sapiens.  
 XX  
 PN MO200192523-A2.  
 XX  
 PD 06-DEC-2001.  
 XX  
 KP 29-MAY-2001; 2001WO-US010836.  
 XX  
 PR 30-MAY-2000; 2000US-0206132P.  
 PR 29-AUG-2000; 2000US-0228716P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shimkels RA, Leach MD;  
 XX  
 DR WPI; 2002-106108/14.  
 DR N-PSDB; ABN25989.  
 XX  
 PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders.  
 XX  
 PS Disclosure; SEQ ID NO 20456; 1037bp; English.  
 XX  
 CC The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see table 1  
 CC in the specification)). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP1500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide

CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage. N.B. The sequence data for this patent did not  
 CC form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pat\_sequences  
 SQ Sequence 139 AA;  
 Query Match 42.9%; Score 6; DB 5; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 ITSVL 12  
 DB 60 ITSVL 65  
 Search completed: December 30, 2004, 15:33:17  
 Job time : 165 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: December 30, 2004, 14:45:25 / Search time 273.065 Seconds  
(without alignments)  
3693.737 Million cell updates/sec

Title: US-10-719-385-2

Perfect score: 9007

Sequence: 1 MRKSKITSVLSFCRSSREL.....PESQEPILQVQAFVHMQR 1753

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

1: uniprot\_02.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8939	99.2	1745	2	014675 mus sapien
2	8482	94.2	1761	2	0620H8
3	8482	94.2	1761	2	BAC97883
4	8270.5	91.8	1639	2	Q723K8
5	6608	73.4	1293	2	Q81W11
6	4093	45.4	853	2	Q9BS12
7	3939	43.7	798	2	Q8C7A1
8	2420	26.9	525	2	Q8R3F1
9	1852	20.6	403	2	Q80U14
10	848	9.4	1056	2	Q7Q9X3
11	829	9.2	190	2	Q89083
12	801.5	8.9	1832	2	Q9V6F6
13	372.5	4.1	1997	2	Q6DRN9
14	368.5	4.1	2025	2	Q8LW11
15	292.5	3.2	2012	1	N205_HUMAN
16	292	3.2	1139	2	Q8S666
17	291.5	3.2	2003	2	Q6PDG0
18	291.5	3.2	2003	2	AAH58729
19	289.5	3.2	2067	2	Q6ZOG1
20	289.5	3.2	2067	2	BAC97903
21	286	3.2	1851	2	Q6P9U5
22	286	3.2	1851	2	Q6P9U5
23	286	3.2	1851	2	AAH60139
24	269	3.0	1641	2	Q6BKR9
25	258.5	2.9	1501	2	Q756X1
26	258.5	2.9	1501	2	AAAS2965
27	258.5	2.9	1670	2	Q6P9U5
28	233.5	2.6	1740	1	N184_SCHPO
29	232	2.6	1710	2	Q6FP09
30	218	2.4	2159	2	Q78Y48
31	217.5	2.4	1818	2	Q6MYH8

32	217.5	2.4	1818	2	CAF32025	CAF32025 aspergill
33	215.5	2.4	561	2	Q9T014	Q9T014 arabidops
34	214	2.4	1655	1	N188_YEAST	P52533 saccharomyc
35	212	2.4	1548	2	Q6C1T5	Q6C1T5 yarrowia 11
36	212	2.4	1606	2	Q6CNP1	Q6CNP1 kluyveromyc
37	205	2.3	1837	2	Q9LUS3	Q9LUS3 arabidops
38	197.5	2.2	1278	2	Q7T153	Q7T153 brachydanio
39	197.5	2.2	1336	2	Q7T152	Q7T152 brachydanio
40	197.5	2.2	2057	2	Q9VWB8	Q9VWB8 drosophila
41	197.5	2.2	2030	2	Q81QV9	Q81QV9 drosophila
42	192.5	2.1	1647	2	P78847	P78847 schizosacch
43	192.5	2.1	1683	1	N192_YEAST	P47054 saccharomyc
44	189	2.1	1248	2	Q9VKY2	Q9VKY2 drosophila
45	184.5	2.0	2670	1	YA05_SCHPO	Q10105 schizosacch
46	179.5	2.0	5280	2	Q7R0B5	Q7R0B5 giardia lam
47	175.5	1.9	1073	2	Q9T013	Q9T013 arabidops
48	175	1.9	2110	2	Q9CAQ9	Q9CAQ9 arabidops
49	171.5	1.9	2104	1	MY33_SCHPO	Q14157 schizosacch
50	169.5	1.9	2443	2	Q96J17	Q96J17 homo sapien
51	169	1.9	1585	2	Q96J17	Q96J17 homo sapien
52	169	1.9	2658	2	Q9X1R5	Q9X1R5 homo sapien
53	167	1.9	4385	2	Q81G62	Q81G62 arabidops
54	166	1.8	2666	2	Q6FW99	Q6FW99 candida gla
55	163.5	1.8	2076	2	Q01485	Q01485 caenorhabdi
56	162.5	1.8	2392	2	Q95001	Q95001 homo sapien
57	162	1.8	2444	2	Q61JM3	Q61JM3 aedes aegyp
58	162	1.8	2615	2	Q7XPP8	Q7XPP8 oryza sativ
59	161	1.8	1858	2	Q7RXM6	Q7RXM6 neurospora
60	160	1.8	1830	1	BP28_ARATH	Q9C844 arabidops
61	160	1.8	1835	2	Q6BMT4	Q6BMT4 debaryomyc
62	159.5	1.8	2598	2	Q7RNO6	Q7RNO6 plasmodium
63	158.5	1.8	2362	2	Q7T160	Q7T160 brachydanio
64	158.5	1.8	3616	2	Q9K6V0	Q9K6V0 gallus gall
65	158	1.8	2036	2	Q7QUA2	Q7QUA2 giardia lam
66	158	1.8	4283	2	Q8UWL7	Q8UWL7 fugu rubrip
67	157.5	1.7	1648	2	Q6CNK8	Q6CNK8 kluyveromyc
68	157.5	1.7	2607	2	Q6L4J5	Q6L4J5 solanum dem
69	157.5	1.7	2607	2	AA70487	AA70487 solanum d
70	157.5	1.7	2736	2	Q6BMQ6	Q6BMQ6 debaryomyc
71	157	1.7	2456	2	Q6F0U9	Q6F0U9 candida gla
72	157	1.7	3859	1	TRAP_HUMAN	Q9Y445 homo sapien
73	157	1.7	5058	2	Q86UQ4	Q86UQ4 homo sapien
74	156	1.7	835	2	Q7KMK9	Q7KMK9 drosophila
75	156	1.7	2605	2	Q81D88	Q81D88 plasmodium
76	156	1.7	2733	2	Q9W246	Q9W246 drosophila
77	155	1.7	2492	1	TALA_DICDI	P54633 dictyostell
78	154.5	1.7	2113	2	Q7Q3X7	Q7Q3X7 anopheles g
79	154.5	1.7	3227	1	PTR1_SCHPO	Q13834 schizosacch
80	154	1.7	2538	2	Q6BXX3	Q6BXX3 debaryomyc
81	154	1.7	2576	2	Q71A77	Q71A77 drosophila
82	154	1.7	2576	2	AAH60444	AAH60444 drosophila
83	154	1.7	3996	2	Q7KTP2	Q7KTP2 drosophila
84	154	1.7	3996	2	AAH10532	AAH10532 drosophila
85	153.5	1.7	2836	2	Q9U3B3	Q9U3B3 caenorhabdi
86	153.5	1.7	2862	2	Q9Y0A2	Q9Y0A2 caenorhabdi
87	153.5	1.7	2882	2	Q9Y0A1	Q9Y0A1 caenorhabdi
88	153	1.7	4568	1	DYHC_CAEEL	Q19020 caenorhabdi
89	153	1.7	5635	2	Q9A9N1	Q9A9N1 leishmania
90	152	1.7	1196	2	Q9A7Z4	Q9A7Z4 linum usita
91	151.5	1.7	1445	2	Q6WAS6	Q6WAS6 brachydanio
92	151.5	1.7	1445	2	AAQ73628	AAQ73628 brachydanio
93	151.5	1.7	1920	2	Q6BNV7	Q6BNV7 debaryomyc
94	151	1.7	3259	1	G0B1_HUMAN	Q14789 homo sapien
95	150.5	1.7	1778	2	Q86XN9	Q86XN9 homo sapien
96	150.5	1.7	8407	2	Q7RTM4	Q7RTM4 homo sapien
97	150	1.7	1403	2	Q80U74	Q80U74 mus musculu
98	150	1.7	2847	2	Q95X74	Q95X74 caenorhabdi
99	150	1.7	2903	2	Q95X75	Q95X75 caenorhabdi
100	149.5	1.7	1334	2	Q9X7S7	Q9X7S7 caenorhabdi
101	149.5	1.7	1998	2	Q8JFV4	Q8JFV4 giardia lam
102	149	1.7	946	2	Q7QVZ0	Q7QVZ0 brachydanio
103	149	1.7	1562	2	Q7QUB1	Q7QUB1 giardia lam
104	148.5	1.6	2675	2	Q92616	Q92616 homo sapien

105 147.5 1.6 1088 1 RP17 HUMAN  
106 147.5 1.6 1088 2 CAC1055  
107 147.5 1.6 2209 1 XNCC HUMAN  
108 147.5 1.6 2241 2 OSMW10  
109 147 1.6 2144 1 BP28 HUMAN  
110 147 1.6 2470 2 O9VK45  
111 147 1.6 2671 2 O754A3  
112 147 1.6 2671 2 AAS53540  
113 147 1.6 2954 2 O42263  
114 147 1.6 6885 1 SNEZ\_HUMAN  
115 146.5 1.6 1170 2 O8MSV6  
116 146.5 1.6 1868 1 YHDO\_YEAST  
117 146.5 1.6 2471 2 O6FNE7  
118 146.5 1.6 2509 2 O786V7  
119 146.5 1.6 2717 2 O7QJ93  
120 146.5 1.6 4829 1 BIR6\_HUMAN  
121 146 1.6 1441 2 O80TF2  
122 146 1.6 2248 1 Y539\_HUMAN  
123 145.5 1.6 1200 2 O9ATZ5  
124 145.5 1.6 1574 2 O8TOD1  
125 145.5 1.6 1629 2 O9BIR5  
126 145.5 1.6 2857 2 O9Y0A0  
127 145.5 1.6 3424 2 O9VT22  
128 145.5 1.6 3479 2 O9VT28  
129 145.5 1.6 3479 2 O9GPT8  
130 145 1.6 1005 1 EVC\_MOUSE  
131 144.5 1.6 1545 2 O8BNO  
132 144.5 1.6 1728 2 O7R067  
133 144.5 1.6 2594 1 7LES\_DROY1  
134 144.5 1.6 5412 2 O7R3N4  
135 144 1.6 1339 2 O7OETO  
136 144 1.6 1358 2 O7XHO5  
137 144 1.6 1358 2 O8W2P4  
138 144 1.6 1407 2 O9WHY0  
139 144 1.6 1600 2 O7S834  
140 144 1.6 1742 2 O9W350  
141 144 1.6 2693 2 O91908  
142 143.5 1.6 611 2 O9D6C1  
143 143.5 1.6 876 2 O8NR86  
144 143.5 1.6 876 2 CARI9871  
145 143.5 1.6 1287 2 O7QXPI  
146 143.5 1.6 2609 2 O7SG70  
147 143.5 1.6 2609 2 CAE76391  
148 143 1.6 2549 1 FRAP\_HUMAN  
149 143 1.6 2564 2 O7YZ64  
150 143 1.6 5162 2 O70LM6

## ALIGNMENTS

RESULT 1  
O14675 PRELIMINARY; PR1: 1745 AA.

ID O14675  
AC O14675;  
DT 01-NOV-1996 (TRENBLREL. 01. Created)  
DT 01-NOV-1996 (TRENBLREL. 01. Last sequence update)  
DT 01-MAR-2004 (TRENBLREL. 26. Last annotation update)  
DE KIAA0169 protein (Fragment).  
GN Name=KIAA0169;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96281124; PubMed=8724849;  
RA Nagase T., Seki N., Ishikawa K., Tanaka A., Nomura N.;  
RT "Prediction of the coding sequences of unidentified human genes. V.  
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by  
RT analysis of cDNA clones from human cell line KG-1.";  
RL DNA Res. 3:17-24(1996).  
RN [2]

RP SEQUENCE FROM N.A.  
RA Ohara O., Nagase T., Kikuno R., Nomura N.,  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; D79991; BA01486.1; -  
DR Genew; HGNC:17859; NUP188.  
FT NON TER 1  
FT NON TER 1  
SQ SEQUENCE 1745 AA; 195697 MW; C774DB49C2F6B871 CRC64;  
Query Match 99.24; Score 8939; DB 2; Length 1745;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1739; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 15 RSSRLTLLIGRALBELSGIEMAKNKRRLLEGISTYTPSPSSAEKTKANKDVASP 74  
DB 7 RSSRLTLLIGRALBELSGIEMAKNKRRLLEGISTYTPSPSSAEKTKANKDVASP 66  
QY 75 LKEGLNISKPLGDEBSQVOLQCYQEDYGRDVKVTLQDERSQALLKADYY 134  
DB 67 LKEGLNISKPLGDEBSQVOLQCYQEDYGRDVKVTLQDERSQALLKADYY 126  
QY 135 EERTCILRCVLLHTYFQDERHPYRVEYACVDKLEKELVSKRQCEELKTEAPTWET 194  
DB 127 EERTCILRCVLLHTYFQDERHPYRVEYACVDKLEKELVSKRQCEELKTEAPTWET 186  
QY 195 HGNMTERQVSRMFWQCLREOSMLLEIFLYAYFEMASDVLVLTQMFQGGSGQTN 254  
DB 187 HGNMTERQVSRMFWQCLREOSMLLEIFLYAYFEMASDVLVLTQMFQGGSGQTN 246  
QY 255 RHLVDETMDFVDRIGFYSALLVEGMDIESLHKCALDDREHQPADGLICODMCLM 314  
DB 247 RHLVDETMDFVDRIGFYSALLVEGMDIESLHKCALDDREHQPADGLICODMCLM 306  
QY 315 LTFGDIPIHAPVLLAWALLNHTINPEETSSVVRKIGSTALQNVFYLTLLQSLAGGN 374  
DB 307 LTFGDIPIHAPVLLAWALLNHTINPEETSSVVRKIGSTALQNVFYLTLLQSLAGGN 366  
QY 375 DCTTSTCMCVGLSFLVLSLHRTGNODDIDRACVTLADPSLELFWGTPFTGGLG 434  
DB 367 DCTTSTCMCVGLSFLVLSLHRTGNODDIDRACVTLADPSLELFWGTPFTGGLG 426  
QY 435 IIDSVCMPFHLSPILLRLALVSGSKPAKKVYSLFDGSPFNLKXKHPDIVSHED 494  
DB 427 IIDSVCMPFHLSPILLRLALVSGSKPAKKVYSLFDGSPFNLKXKHPDIVSHED 486  
QY 495 GTLMRRQTPFLVPLGQTNLRIPOGTVGQMLDRAVLRWEXYSWTLFTCEIEML 554  
DB 487 GTLMRRQTPFLVPLGQTNLRIPOGTVGQMLDRAVLRWEXYSWTLFTCEIEML 546  
QY 555 HVTSTADVLOHCRQVRPIIDLVRKVTSTDSIADCLPISRIYMLQRLTTYISPPDV 614  
DB 547 HVTSTADVLOHCRQVRPIIDLVRKVTSTDSIADCLPISRIYMLQRLTTYISPPDV 606  
QY 615 IASCVNCLTVLAARNPAKVTDLRHGTFLPFVAPVSSLSQMSABGMAGYGNLMS 674  
DB 607 IASCVNCLTVLAARNPAKVTDLRHGTFLPFVAPVSSLSQMSABGMAGYGNLMS 666  
QY 675 EOPQSGEVTVIAFLRIITLVKQIGSTQSGILVPCVMFVLKXMLPSYHWRVNSHVR 734  
DB 667 EOPQSGEVTVIAFLRIITLVKQIGSTQSGILVPCVMFVLKXMLPSYHWRVNSHVR 726  
QY 735 QIGCLILHIALINCHERDLHSSHTPSLOFCTISLATYBAGQVYINMGVDTIDM 794  
DB 727 QIGCLILHIALINCHERDLHSSHTPSLOFCTISLATYBAGQVYINMGVDTIDM 786  
QY 795 VMAAPRSDAGSGGQGLIKTVKLAFASTVNVIRLKPSSNVSPLEQALSOHAGGN 854  
DB 787 VMAAPRSDAGSGGQGLIKTVKLAFASTVNVIRLKPSSNVSPLEQALSOHAGGN 846  
QY 855 LIAVLAKTYTHKIDPALPRLAQQLRLATVAEMSVYACVGNDAARLDAFLTRLOSKIB 914  
DB 847 LIAVLAKTYTHKIDPALPRLAQQLRLATVAEMSVYACVGNDAARLDAFLTRLOSKIB 906

```

QY 915 DMRIKMIIEFLTAVAVETOPGLIELFLNLBYKDGSDGSKPSIGWMSCLAHVLEIDSDQ 974
Db 907 DMRIKMIIEFLTAVAVETOPGLIELFLNLBYKDGSDGSKPSIGWMSCLAHVLEIDSDQ 966
QY 975 QDRYWCPEPLHRAAIAFLHALMODRSDSAMLVLRKPKFWMNTLSPLEFGLSPSSSETSEP 1034
Db 967 QDRYWCPEPLHRAAIAFLHALMODRSDSAMLVLRKPKFWMNTLSPLEFGLSPSSSETSEP 1026
QY 1035 SILETCALIMKICLEIETVYVYKSGSDSLKOTLKPKFSIEKPAVWSGVYKSLAVHVAETE 1094
Db 1027 SILETCALIMKICLEIETVYVYKSGSDSLKOTLKPKFSIEKPAVWSGVYKSLAVHVAETE 1086
QY 1095 GSSCTSLLEYOMLSAARMMLIIATTADIMHLDSDVSRROLPLDVLDTGATALLVPASV 1154
Db 1087 GSSCTSLLEYOMLSAARMMLIIATTADIMHLDSDVSRROLPLDVLDTGATALLVPASV 1146
QY 1155 NCURLGSMKCTLLIILKQMRBELGSVDEILGPLETEILEGYLOADQOLMEKTKAVPSAF 1214
Db 1147 NCURLGSMKCTLLIILKQMRBELGSVDEILGPLETEILEGYLOADQOLMEKTKAVPSAF 1206
QY 1215 ITVLQMKEMKVSIDIPOYSQVLAVNCETIQEBVIALFDQTRHSLAGSATEDKDSMETDDC 1274
Db 1207 ITVLQMKEMKVSIDIPOYSQVLAVNCETIQEBVIALFDQTRHSLAGSATEDKDSMETDDC 1266
QY 1275 SRSRHROROGCVLGLHLAKELCVDESDSWLOVRRLPILPILTLTEVSLRMKNTL 1334
Db 1267 SRSRHROROGCVLGLHLAKELCVDESDSWLOVRRLPILPILTLTEVSLRMKNTL 1336
QY 1335 HFEATLHLTLTLARTOOGATAVAGAGITQSIGCLPLSVYQSTNGTQTPSASRKSIDA 1394
Db 1327 HFEATLHLTLTLARTOOGATAVAGAGITQSIGCLPLSVYQSTNGTQTPSASRKSIDA 1386
QY 1395 PSMGVRVRLSNLSMEQLIKTLRYNPLPEALDPVGVHQRERTLCQNAVTVQSLACLEAD 1454
Db 1387 PSMGVRVRLSNLSMEQLIKTLRYNPLPEALDPVGVHQRERTLCQNAVTVQSLACLEAD 1446
QY 1455 HTVGFIIQLSNFMKEMHFLPOLMRDIOVNGYLCQACTSLHSHRKMLOHLOKNGDGL 1514
Db 1447 HTVGFIIQLSNFMKEMHFLPOLMRDIOVNGYLCQACTSLHSHRKMLOHLOKNGDGL 1506
QY 1515 PSAVAGORVPPSAASAAPSSSKOPADTEASEOALHTVOYGLKILSKTLAALRHPT 1574
Db 1507 PSAVAGORVPPSAASAAPSSSKOPADTEASEOALHTVOYGLKILSKTLAALRHPT 1566
QY 1575 DVCQILLDQSLDLAEYNFLPALSTFTPTFDSVAPSPFGTLLATYVVALNMLGELDKKEP 1634
Db 1567 DVCQILLDQSLDLAEYNFLPALSTFTPTFDSVAPSPFGTLLATYVVALNMLGELDKKEP 1626
QY 1635 LTOAVGLSTQABGRITLKSLLMFTMENCYLLISQANRYLDPVAVHPDKQRMKQELSSSE 1694
Db 1627 LTOAVGLSTQABGRITLKSLLMFTMENCYLLISQANRYLDPVAVHPDKQRMKQELSSSE 1686
QY 1695 LSTLSSLSRFRGAPSSPATGVLPSPQGSTLSKSPSOSPELIDLOVAFVPHMOR 1753
Db 1687 LSTLSSLSRFRGAPSSPATGVLPSPQGSTLSKSPSOSPELIDLOVAFVPHMOR 1745

```

## RESULT 2

```

06ZOH8 PRELIMINARY; PRT; 1761 AA.
AC 06ZOH8;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE MRLA0169 protein (fragment).
GN Name=MRLA0169;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic tail;

```

```

RX PubMed=14621295;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hirooka S.,
RA Saga Y., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT III. the complete nucleotide sequences of 500 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:167-180(2003).
DR EMBL; AK129073; BAC97883.1; -.
DR InterPro; IPR008938; ARM.
FT NON TER 1
SQ SEQUENCE 1761 AA; 196893 MW; 20CE3DC250EBED93 CRC64;

Query Match 94.2%; Score 8482; DB 2; Length 1761;
Beet Local Similarity 93.9%; Pred. No. 0;
Matches 1643; Conservative 51; Mismatches 45; Indels 10; Gaps 1;

QY 15 RSSRELWTIILGSRALRELSQLIBALNKHNRRLLEGSYKPPSPSSAEKYANKDVASP 74
Db 13 RSSRELWTIILGSRALRELSQLIBALNKHNRRLLEGSYKPPSPSSAEKYANKDVASP 72
QY 75 LKELGLRISKPLGDBESQVQLQCYIQEDYRGTRDSVKTVLQDEROSQALLIKIADYY 134
Db 73 LKELGLRISKPLGDBESQVQLQCYIQEDYRGTRDSVKTVLQDEROSQALLIKIADYY 132
QY 135 EERTCIILCTVHLTLTYQDERHPRVAVYADCVDLKEKELYSKROQPEELYTEAPMET 194
Db 133 EERTCIILCTVHLTLTYQDERHPRVAVYADCVDLKEKELYSKROQPEELYTEAPMET 192
QY 195 HGNLATEROVSRLVOCLEBSMLLEIIFLYAFAENAPSDLVLTVMFEQSGSRQTN 254
Db 193 HGNLATEROVSRLVOCLEBSMLLEIIFLYAFAENAPSDLVLTVMFEQSGSRQTN 252
QY 255 RHLVDETMDFPVDRIGFYSALLVVBGMDISLHKCALDDBREHLQPAQDGLICQMDCLM 314
Db 253 RHLVDETMDFPVDRIGFYSALLVVBGMDISLHKCALDDBREHLQPAQDGLICQMDCLM 312
QY 315 LTFGDIPIHAPVLLAAMLLKHTLNPEBTSYVYKIGGTATQIANFYOTLTLTSLASGNN 374
Db 313 LTFGDIPIHAPVLLAAMLLKHTLNPEBTSYVYKIGGTATQIANFYOTLTLTSLASGNN 372
QY 375 DCTTSTACMCVYGLSLFVLSLSELTGNODIIDTACEVLAADSPLELFWGTPSGIG 434
Db 373 DCTTSTACMCVYGLSLFVLSLSELTGNODIIDTACEVLAADSPLELFWGTPSGIG 432
QY 435 IILDVCGMPHLLSPILLQILRALVSGSKTAKVYSFLDKNSFYNELYKHNPDIYSHED 494
Db 433 IILDVCGMPHLLSPILLQILRALVSGSKTAKVYSFLDKNSFYNELYKHNPDIYSHED 492
QY 495 GTLMRRQTPKLLPYLGGQTNLRIPQGTGVGYMLDDRAVLYRWEXYSYSWTLFTCEIEMTL 554
Db 493 GTLMRRQTPKLLPYLGGQTNLRIPQGTGVGYMLDDRAVLYRWEXYSYSWTLFTCEIEMTL 552
QY 555 HVYSTADVIOHCORVKRVIDLVHNVISTDISINDCLLPITSRIYMLQRLTVVISPVDV 614
Db 553 HVYSTADVIOHCORVKRVIDLVHNVISTDISINDCLLPITSRIYMLQRLTVVISPVDV 612
QY 615 IASCNCLITVLAANPAKWTDLRHGTFLPFVAHPVSSLSQMTSABGNAGYGNLIMS 674
Db 613 IASCNCLITVLAANPAKWTDLRHGTFLPFVAHPVSSLSQMTSABGNAGYGNLIMS 672
QY 675 EOPQGEYGVITIAFLRITTVYKGLSGTQSGVLPVCMFYLKEMLPYHKKRNVSHGVR 734
Db 673 EOPQGEYGVITIAFLRITTVYKGLSGTQSGVLPVCMFYLKEMLPYHKKRNVSHGVR 732
QY 735 QIGCLTIELHAIINLCHETDILSHSPSLQPLCTIGSLATBAGQVYINIMIGIVDTIDM 794
Db 733 QIGCLTIELHAIINLCHETDILSHSPSLQPLCTIGSLATBAGQVYINIMIGIVDTIDM 792
QY 795 VMAAQRPSDABEGGQGLIKTVKLAFSVTNNVIRLKPSSNVVSPLEQALSQHGAGNN 854
Db 793 VMAAQRPSDABEGGQGLIKTVKLAFSVTNNVIRLKPSSNVVSPLEQALSQHGAGNN 852

```

```

QY 855 LIAVLAKYIYKHDPALPRLAIOILKRLATVA PMSYVACLGNDAAIRDPILRLQSKTE 914
DB 853 LIAVLAKYIYKHDPALPRLAIOILKRLATVA PMSYVACLGNDAAIRDPILRLQSKTE 912
QY 915 DMRIRKMIIEPLTVA VETOPGILIEPLNLEVDGSGSKFSLGMSCLHVALELIDSOQ 974
DB 913 DMRIRKMIIEPLTVA VETOPGILIEPLNLEVDGSGSKFSLGMSCLHVALELIDSOQ 972
QY 975 QDRYVCPLLRHAAAFALHAAQDRRDSAMLYRTKPKWENUTSPFLFTLSPSETSEP 1034
DB 973 QDRYVCPLLRHAAAFALHAAQDRRDSAMLYRTKPKWENUTSPFLFTLSPSETSEP 1032
QY 1035 SLLETICALIMKICIEIYVVGASLDOSKDTLKFSIKRPAVYSGYKSLAVVAE 1094
DB 1033 SVLEFCALIMKICIEIYVVGASLDOSKDTLKFSIKRPAVYSGYKSLAVVAE 1092
QY 1095 GSSCTSLLEYOMLVASAMRMLIIATTHADIMHITDSVVRQDFLDVDTKALLVPASV 1154
DB 1093 GSSCTSLLEYOMLVASAMRMLIIATTHADIMHITDSVVRQDFLDVDTKALLVPASV 1152
QY 1155 NCLRIGSKKCTLLILRLQMKELGSDVEIIGPTEILEGVQADQOLMEKTKAVFSAF 1214
DB 1153 NCLRIGSKKCTLLILRLQMKELGSDVEIIGPTEILEGVQADQOLMEKTKAVFSAF 1212
QY 1215 ITVLQMKEMKYSDIPQYSOLVANCETLOEVIYALFDOTRHSALGSAATEDDSMETDC 1274
DB 1213 ITVLQMKELRVQDIQYISQVLVANCETLOEVIYALFDOTRHSALGSAATEDDSMETDC 1272
QY 1275 SRSRRRDQDGVCVGLHAAKELCEVDBDGSMLQVTRRLPILPTLLTLEVSLEMKONT 1334
DB 1273 PRPRHKQDQDGVCVGLHAAKELCEVDBDGSMLQVTRRLPILPTLLTLEVSLEMKONT 1332
QY 1335 HPTETALHLLTLAATQOGATVAVAGAGITQSCIEPLLSVYQUSNTGCTQTSASKSIDA 1394
DB 1333 HPTETALHLLTLAATQOGATVAVAGAGITQSCIEPLLSVYQUSNTGCTQTSASKSIDA 1392
QY 1395 PSWPGVYVLSMSLEQLKTLRYNPLPALDFVGHQERTLOCLNAVRYTVOGLACLEAD 1454
DB 1393 PSWPGVYVLSMSLEQLKTLRYNPLPALDFVGHQERTLOCLNAVRYTVOGLACLEAD 1452
QY 1455 HTVGPILQLSNPKEMHFLPOLMMDIQVNLGYLCOACTSLHSHKMLQHTLQNNNGDGL 1514
DB 1453 HTVGPILQLSNPKEMHFLPOLMMDIQVNLGYLCOACTSLHSHKMLQHTLQNNNGDGL 1512
QY 1515 PSAVAVQVRBPESA-----ASAPSSSKOPAAATEASBOQALHTVGYGLKILSK 1564
DB 1513 PSAVAVQVRBPESATTTTATATAPAGCSSKOPADTEASEQALHTVGYGLKILSK 1572
QY 1565 TLALRHPTPVCQIILDDQSLDLAEVNFALSFPTTFDSEVAPSPFTLLATVVALNM 1624
DB 1573 TLALRHPTPVCQIILDDQSLDLAEVNFALSFPTTFDSEVAPSPFTLLATVVALNM 1632
QY 1625 LGEIDKKEPELTOAVGSTQAEGRITLKSILMFTMENC FYLLISQAKRYLDPVHPBDK 1664
DB 1633 LGEIDKKEPELTOAVGSTQAEGRITLKSILMFTMENC FYLLISQAKRYLDPVHPBDK 1692
QY 1685 QRMKQELSSSELSTLSSLSRYFRGAPSSPATGVLPSPQGSTLSKSPESOBPLIOV 1744
DB 1693 QRMKQELSSSELSTLSSLSRYFRGAPSSPATGVLPSPQGSTLSKSPESOBPLIOV 1752
QY 1745 QAFVHRMQR 1753
DB 1753 QAFVHRMQR 1761

```

```

RESULT 3
BAC97883 PRELIMINARY; PRF, 1761 AA.
AC BAC97883;
DT 02-MAR-2004 (Tremblrel. 27, Created)
DT 02-MAR-2004 (Tremblrel. 27, Last sequence update)
DT 02-MAR-2004 (Tremblrel. 27, Last annotation update)
DE KIAA0169 protein (Fragment).

```

```

GN KIAA0169.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISUB=Embryonic tail;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hirooka S.,
RA Suga Y., Nagase T., Ohara O., Koga H.;
RT "Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene:
RT Iii. The Complete Nucleotide Sequences of 500 Mouse KIAA-homologue
RT cDNAs Identified by Screening of Terminal Sequences of cDNA Clones
RT Randomly Sampled from Size-fractionated Libraries.";
RL DNA Res. 10:167-180(2003).
DR EMBL; AK129073; BAC97883.1; -.
FT NON TER
SQ SEQUENCE 1761 AA; 196893 MW; 20CE3DC250EBBD93 CRC64;

Query Match 94.2%; Score 8482; DB 2; Length 1761;
Best Local Similarity 93.98; Pred. No. 0;
Matches 1643; Conservative 51; Mismatches 45; Indels 10; Gaps 1;

QY 15 RSSRELMTIILGRALRELQIEBALNKHMRRLIEGLSYKPPSPSSAKRYKANKDVASP 74
DB 13 RSSRELMTIILGRALRELQIEBALNKHMRRLIEGLSYKPPSPSSAKRYKANKDVASP 72
QY 75 LKEIGLRISKEPLGDBEQSVQLQCYLOEDYRGTRDSKVTVLQDEROSQALILKTDVYY 134
DB 73 LKEIGLRISKEPLGDBEQSVQLQCYLOEDYRGTRDSKVTVLQDEROSQALILKTDVYY 132
QY 135 EERTCILRCVHLITTYQDERHPRVAVYADCVDLKEELYSKTRQOEELKYETAPEWET 194
DB 133 EERTCILRCVHLITTYQDERHPRVAVYADCVDLKEELYSKTRQOEELKYETAPEWET 192
QY 195 HGNLMTROYSRMFWQCLREOSMLBIIIFYVYFEMAPSDLVLTVMFKQSGSROT 254
DB 193 HGNLMTROYSRMFWQCLREOSMLBIIIFYVYFEMAPSDLVLTVMFKQSGSROT 252
QY 255 RHVDETMDPVDRIQFSAALIVEGMDIESLHKCALDDBREILHQAQDGLICQMDCLM 314
DB 253 RHVDETMDPVDRIQFSAALIVEGMDIESLHKCALDDBREILHQAQDGLICQMDCLM 312
QY 315 LTPEDTIPHAAPVLLAAMLLRHTLNPEBTSVYKIGSTAIQANVQVLTLLQSLAGCN 374
DB 313 LTPEDTIPHAAPVLLAAMLLRHTLNPEBTSVYKIGSTAIQANVQVLTLLQSLAGCN 372
QY 375 DCTTSTACMCVYGLISFVLSLEHTLGNQODIIDTACEVYADPSLPFLFMGTEPTSGLG 434
DB 373 DCTTSTACMCVYGLISFVLSLEHTLGNQODIIDTACEVYADPSLPFLFMGTEPTSGLG 432
QY 435 IILDVCGMFPHLISPLQLLRALVSGKSTAKKVSFLDKMSFYNELYKRPDVISHD 494
DB 433 IILDVCGMFPHLISPLQLLRALVSGKSTAKKVSFLDKMSFYNELYKRPDVISHD 492
QY 495 GTLMRQTPCLLPLGGQNTLRIPQGTGQVMDLDRAYLVKRWYSISWSWLPFCEIMLL 554
DB 493 GTLMRQTPCLLPLGGQNTLRIPQGTGQVMDLDRAYLVKRWYSISWSWLPFCEIMLL 552
QY 555 HVTSTADVIOHCORVPIIDLVKRVISTDLSIADCLLPIYSRIYMLQRLTTYVSPVDV 614
DB 553 HVTSTADVIOHCORVPIIDLVKRVISTDLSIADCLLPIYSRIYMLQRLTTYVSPVDV 612
QY 615 IASCVNCLTVLAARNPAKVTWTLRHGTPLPFAVHPVSSLSQMSIABGMNAGYGNILMS 674
DB 613 IASCVNCLTVLAARNPAKVTWTLRHGTPLPFAVHPVSSLSQMSIABGMNAGYGNILMS 672
QY 675 EOPQGEYGTIAIRLITLVYQGLSTQSGVLPVCMFTLKMLSYNHRNYSIGVRE 734
DB 673 EOPQGEYGTIAIRLITLVYQGLSTQSGVLPVCMFTLKMLSYNHRNYSIGVRE 732
QY 735 QIGCLILELHAINTCHETDLSSHTPSIQFCISLAVTEAGQVYINIMGIGVPTIDM 794
DB 733 QIGCLILELHAINTCHETDLSSHTPSIQFCISLAVTEAGQVYINIMGIGVPTIDM 792

```

```

Db 733 LIGCLILELHAIHNLCOETELSHSHTSPSLCISLAVTEAGGTVISIMGIVDTIDM 792
Qy 795 VMAAPRSDGAEAGGCGGOLLITTVKLASVTNNVRLKPPSNVSPLEBALSOQHAGNN 854
Db 793 VMAAPRSDGAEAGGCGGOLLITTVKLASVTNNVRLKPPSNVSPLEBALSOQHAGNN 852
Qy 885 LIAVAKTIYHNDALPRLAIQLKRLATVAPMSVYACLGDAAIRAPFRLQSKIE 914
Db 853 LIAVAKTIYHNDALPRLAIQLKRLATVAPMSVYACLGDAAIRAPFRLQSKIE 912
Qy 915 DMRIKMIIEFLTVAVEFQPIELFLNLEVDGSDGSEFSLGMSCLHVLIEDID9Q 974
Db 913 DMRIKMIIEFLTVAVEFQPIELFLNLEVDGSDGSEFSLGMSCLHVLIEDID9Q 972
Qy 975 QDRYWCPLLRRAIAFLHAWODRRDSAMLVLRTPKRWENUTSPFLFTLSPPSTSP 1034
Db 973 QDRYWCPLLRRAIAFLHAWODRRDSAMLVLRTPKRWENUTSPFLFTLSPPSTSP 1032
Qy 1035 SILETCALIMKICIEIYVVGSLDOSLKOITKKPSIEKRPVMSGYKSLAVHAE 1094
Db 1033 SVLETCALIMKICIEIYVVGSLDOSLKOITKKPSIEKRPVMSGYKSLAVHAE 1092
Qy 1095 GSSCTSLSEYOMLVASAMRLIATTHADIMHLDTSVVARQLFDVLDGTALLVPSV 1154
Db 1093 GSSCTSLSEYOMLVASAMRLIATTHADIMHLDTSVVARQLFDVLDGTALLVPSV 1152
Qy 1155 NCLRIGSNKCTLLILLRWKKEIDSVDEILDPLEILGVTQADQOLMEKTKAKYPSAF 1214
Db 1153 NCLRIGSNKCTLLILLRWKKEIDSVDEILDPLEILGVTQADQOLMEKTKAKYPSAF 1212
Qy 1215 ITTVLQMKMKVSDIOYSQLVNNCELOEVIYALPDORHSLAAGSATFEDSDMETDC 1274
Db 1213 ITTVLQMKMKVSDIOYSQLVNNCELOEVIYALPDORHSLAAGSATFEDSDMETDC 1272
Qy 1275 SRSRARDODGVCVGLHAKELCEVEDSDMLQVTRRLPIPLTLTLEVSLMKOML 1334
Db 1273 SRSRARDODGVCVGLHAKELCEVEDSDMLQVTRRLPIPLTLTLEVSLMKOML 1332
Qy 1335 HPTBATLHLLTLARTQOGATAVAGAGITQOSICPLSLSYQLSTNGTAOTPSAKSKSLDA 1394
Db 1333 HPTBATLHLLTLARTQOGATAVAGAGITQOSICPLSLSYQLSTNGTAOTPSAKSKSLDA 1392
Qy 1395 PSMPEVYRLSMGLMEQLKTLRYNPLPALDPVGNQERTLOCLNAVTRVQSLACLEAD 1454
Db 1393 PSMPEVYRLSMGLMEQLKTLRYNPLPALDPVGNQERTLOCLNAVTRVQSLACLEAD 1452
Qy 1455 HTVGFILQISNFMKEHFFHLPOLMRDIOVNLGYLCOACTSLHSRMLQHYLONKNGDGL 1514
Db 1453 HTVGFILQISNFMKEHFFHLPOLMRDIOVNLGYLCOACTSLHSRMLQHYLONKNGDGL 1512
Qy 1515 PSAVAQVQRPSPA-----ASAPSSSKOPADTEASEBOALHTVOYGLKILSK 1564
Db 1513 PSAVAQVQRPSPA-----ASAPSSSKOPADTEASEBOALHTVOYGLKILSK 1562
Qy 1565 TLAAALRHFTPDVCOILLDOSLDAEYNPLPALSFTTPTDSBVAPEFGTLLATVNALNM 1624
Db 1573 TLAAALRHFTPDVCOILLDOSLDAEYNPLPALSFTTPTDSBVAPEFGTLLATVNALNM 1632
Qy 1635 LGEIDKKKEPLTOAVGLSTQABSTRKLSLMTMNCFYLLISQAMRYLADPAVPRDK 1684
Db 1633 LGEIDKKKEPLTOAVGLSTQABSTRKLSLMTMNCFYLLISQAMRYLADPAVPRDK 1692
Qy 1685 QMKOBSLSELSTLSSLSRYFRGAPSSPATGVLSPQKSTSLSKAPESQEPILQV 1744
Db 1693 QMKOBSLSELSTLSSLSRYFRGAPSSPATGVLSPQKSTSLSKAPESQEPILQV 1752
Qy 1745 QAFVRRMOR 1753
Db 1753 QAFVRRMOR 1761

```

---

```

ID 0723K8 PRELIMINARY; PRT; 1639 AA.
AC 0723K8;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein DKFZp686l1653 (fragment).
GN Name=DKFZp686l1653;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC Tissue=Human endometrium;
RA Nabut R., Heubner D., Mewes H.W., Weil B., Amlid C., Osanger A.,
RA Fodor G., Han M., Miesman S.,
RL Submitted (JUN-2003) to the EMBL/Genbank/DBJ databases.
DR EMBL, BX537774; CAD97835.1; -.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 1639 AA; 182478 MW; B58334E14B217B92 CRC64;

Query Match 91.8%; Score 8270.5; DB 2; Length 1639;
Best Local Similarity 93.5%; Pred. No. 0;
Matches 1626; Conservative 1; Mismatches 1; Indels 111; Gaps 1;

15 RSSRLWTLILGRSLRLRLSGLBRLNGHWRLLRGLSYKPPSSAKRYANDVASP 74
12 RSSRLWTLILGRSLRLRLSGLBRLNGHWRLLRGLSYKPPSSAKRYANDVASP 71
75 LKELGLRISKFLGDLDEOSVOLLOCYLQEDYKGRDTSYKTVLQDERQSOALLIKIADYY 134
72 LKELGLRISKFLGDLDEOSVOLLOCYLQEDYKGRDTSYKTVLQD----- 115
135 EERTCILRCVHLHTTYFODERHPRVRYADCVDKLEKELVSKIRQOFSELYTEAPTMT 194
116 ----- 115
135 HGNLTERQSRNPFQCLAREQSMLEIIFLYAYREMAPSDLVLTMKPKGQPSRQTN 254
116 -----EMAPSDLVLTMKPKGQPSRQTN 140
255 RHLVDETMDPFVDRIGYFSALLIVEGMDIESLHKCALDRRLRHQFAODGLICODMDCLM 314
141 RHLVETMDPFVDRIGYFSALLIVEGMDIESLHKCALDRRLRHQFAODGLICODMDCLM 200
315 LTFGDIPIHAPVTLAMALRHNLNPEERTSSVVRKIGGTALQUNVEOYLTRLLOSLAGSN 374
201 LTFGDIPIHAPVTLAMALRHNLNPEERTSSVVRKIGGTALQUNVEOYLTRLLOSLAGSN 260
375 DCTTSTACMCVYGGLISFVLTSLRHTLGNQDIIITPACGVLADPSLPFLFNGTPTPSGLG 434
261 DCTTSTACMCVYGGLISFVLTSLRHTLGNQDIIITPACGVLADPSLPFLFNGTPTPSGLG 320
435 ILLDVCVGFPHLSPLIQLALBALVSGKSTAKKYSPFLDKMSFYNELYGKRDVISHED 494
321 ILLDVCVGFPHLSPLIQLALBALVSGKSTAKKYSPFLDKMSFYNELYGKRDVISHED 380
495 GTLAMRQRPKLLYPLGGQTNLRIPOGTQVQVMDRAVLYVREYSYSWTFLTCETEMLL 554
381 GTLAMRQRPKLLYPLGGQTNLRIPOGTQVQVMDRAVLYVREYSYSWTFLTCETEMLL 440
555 HVSSTADVIOHCORVKEPIIDLVHKTISTDLSTADCLPPTSNIYMLQRLTVISPPDV 614
441 HVSSTADVIOHCORVKEPIIDLVHKTISTDLSTADCLPPTSNIYMLQRLTVISPPDV 500
615 IASCVNCLTVLAARPAKWTDLRHTGFLPFAHPVPSLSQMSISAGNAGGYNLKMS 674
501 IASCVNCLTVLAARPAKWTDLRHTGFLPFAHPVPSLSQMSISAGNAGGYNLKMS 560
675 EOPQGEYGTIAFLRLITLVKQGLSSTQGLVPCVMEVLKEMLPSTYKMRVNSHGV 734
561 EOPQGEYGTIAFLRLITLVKQGLSSTQGLVPCVMEVLKEMLPSTYKMRVNSHGV 620

```

```

QY 735 QIGGLIIEIHLAIINLCHETDISHSHPSIOFLICISLAATAGQTYINIMGIGVDTIDM 794
DB 621 QIGGLIIEIHLAIINLCHETDISHSHPSIOFLICISLAATAGQTYINIMGIGVDTIDM 680
QY 795 VMAAQRSDAEGGQGGQGLIKTYKLAFSVTNNVIRLKPSPNVVSPLQALSOCHAGANN 854
DB 681 VMAAQRSDAEGGQGGQGLIKTYKLAFSVTNNVIRLKPSPNVVSPLQALSOCHAGANN 740
QY 855 LIAVLAKYIYKHDPALPRLAIQLKLAATYAPMSVYACIGNDAAIRDAFLTRLOSKIE 914
DB 741 LIAVLAKYIYKHDPALPRLAIQLKLAATYAPMSVYACIGNDAAIRDAFLTRLOSKIE 800
QY 915 DMKIKWIEFLIYAVETOPGLIELFNLBYKQSDSDSKPSICMMSCLHAYELIDSOQ 974
DB 801 DMKIKWIEFLIYAVETOPGLIELFNLBYKQSDSDSKPSICMMSCLHAYELIDSOQ 860
QY 975 QDRYWCPELHRAAIAFLHLMODRDSAMLVIRTKPEWENTLSPULGTLSPSETSEP 1034
DB 861 QDRYWCPELHRAAIAFLHLMODRDSAMLVIRTKPEWENTLSPULGTLSPSETSEP 920
QY 1035 SIETCALIMKIIICLETYYVYKSLDQSLDKKSEIEKRFAYWSGYKSLAVHVAET 1094
DB 921 SIETCALIMKIIICLETYYVYKSLDQSLDKKSEIEKRFAYWSGYKSLAVHVAET 980
QY 1095 GSSCTSLLEYQMLVSARMLLIIATTHADIMHLLTDSVVRQLFADVDGKALLVPAV 1154
DB 981 GSSCTSLLEYQMLVSARMLLIIATTHADIMHLLTDSVVRQLFADVDGKALLVPAV 1040
QY 1155 NCIRLGSMTCTLLIILRQKRELGSDVIELGLPTEILLEGVLOADQOLMEKTYKVSAP 1214
DB 1041 NCIRLGSMTCTLLIILRQKRELGSDVIELGLPTEILLEGVLOADQOLMEKTYKVSAP 1100
QY 1215 ITVLQKMKEMVSDIIPQYSLVAVNCETLOEBEVALPDQTHSLALSGATDCKSMETD 1274
DB 1101 ITVLQKMKEMVSDIIPQYSLVAVNCETLOEBEVALPDQTHSLALSGATDCKSMETD 1160
QY 1275 SRSRHRDQDGVCLGLHLAKELCEVEDDSDMLQVTRRLPIPTLLITLVEISLRKONL 1334
DB 1161 SRSRHRDQDGVCLGLHLAKELCEVEDDSDMLQVTRRLPIPTLLITLVEISLRKONL 1220
QY 1335 HFTBATLHLLLTARTQOGATAVAGAGITQSIGLPILSYVQSLSTNGTAQTPSARSLDA 1394
DB 1221 HFTBATLHLLLTARTQOGATAVAGAGITQSIGLPILSYVQSLSTNGTAQTPSARSLDA 1280
QY 1395 PSMPGYRLSMISMEQLKTLRYNPLPEALDFYGVCHOERTLOCLNAVTRVQSLACLEAD 1454
DB 1281 PSMPGYRLSMISMEQLKTLRYNPLPEALDFYGVCHOERTLOCLNAVTRVQSLACLEAD 1340
QY 1455 HTVGFILQSLNFKEMHFLPOLMRDIOVNLGYLCOACTSLHARRKOLQHYLONKQDGL 1514
DB 1341 HTVGFILQSLNFKEMHFLPOLMRDIOVNLGYLCOACTSLHARRKOLQHYLONKQDGL 1400
QY 1515 PSVAVORVPRPSASAPSSSKQAPADTEASQALHTVOYGLKILSKTALAHFPP 1574
DB 1401 PSVAVORVPRPSASAPSSSKQAPADTEASQALHTVOYGLKILSKTALAHFPP 1460
QY 1575 DVCQILDDSLDAEYNFLPALSFPTPTDSEVABPFGTLAVVNALMMLGELDKKBP 1634
DB 1461 DVCQILDDSLDAEYNFLPALSFPTPTDSEVABPFGTLAVVNALMMLGELDKKBP 1520
QY 1635 LTVQAVGLSTQAGSTRILKSLMFTWENCFLILISQMYRLDPAVHPRDKORQKOLSE 1694
DB 1521 LTVQAVGLSTQAGSTRILKSLMFTWENCFLILISQMYRLDPAVHPRDKORQKOLSE 1580
QY 1695 LSTLLSLSRYRRGAPSSPANGVLPSPQKSTSLSKASPESEGLIQLVQAFVRMOR 1753
DB 1581 LSTLLSLSRYRRGAPSSPANGVLPSPQKSTSLSKASPESEGLIQLVQAFVRMOR 1639

```

RESULT 5  
Q81WF1 PRELIMINARY; PRT; 1293 AA.

```

AC Q81WF1;
DT 01-MAR-2003 (Tremblé). 23, Created
DT 01-MAR-2003 (Tremblé). 23, Last sequence update
DT 01-MAR-2004 (Tremblé). 26, Last annotation update
DE NUP188 protein (Fragment).
GN Name=NUP188;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Krausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heigh P.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Fange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Millany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Viallon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywicki M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Strausberg R.;
RA Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC040352; AAK0352.1; -.
DR InterPro; IPR008938; ARM.
FT NON TER
SQ SEQUENCE 1293 AA; 144063 MW; F0D1B7BE69DC1791 CRC64;
Query Match 73.4%; Score 6608; DB 2; Length 1293;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1292; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 461 GKSTAKKYSPFDRKSPFNELYKHKPHDVISHEGDTLMRQPKLLYPLGQTNLRIPG 520
DB 1 GKSTAKKYSPFDRKSPFNELYKHKPHDVISHEGDTLMRQPKLLYPLGQTNLRIPG 60
QY 521 TTGQVMDLDRAYLVWMEYSYSSWTLFTCEIEKLHVSSTADYIQHQRVPIIDLVHXY 580
DB 61 TTGQVMDLDRAYLVWMEYSYSSWTLFTCEIEKLHVSSTADYIQHQRVPIIDLVHXY 120
QY 581 STDLSIADCLPITSRIYMLQRLTTVISPPDVASCVNCLTVAAANPAVMDLRHT 640
DB 121 STDLSIADCLPITSRIYMLQRLTTVISPPDVASCVNCLTVAAANPAVMDLRHT 180
QY 641 GFLPFAHVPVSLSGMISABGNAGYGNLMSBQPGYGVYTAFLRLITLVKQGLG 700
DB 181 GFLPFAHVPVSLSGMISABGNAGYGNLMSBQPGYGVYTAFLRLITLVKQGLG 240
QY 701 STQSGALVPCVWFVLEKLPSTYHKWRVNSHGVRQIGCLIELIHLAIINLCHETDLSSH 760
DB 241 STQSGALVPCVWFVLEKLPSTYHKWRVNSHGVRQIGCLIELIHLAIINLCHETDLSSH 300
QY 761 TPSLOFLICISLAATAGQTYINIMGIGVDTIDMMAAQRSDAEGGQGGQGLIKTYKL 820
DB 301 TPSLOFLICISLAATAGQTYINIMGIGVDTIDMMAAQRSDAEGGQGGQGLIKTYKL 360
QY 821 AFSVTNNVIRLKPSPNVVSPLQALSOCHAGANNLIAVLAKYIYKHDPALPRLAIQLK 880

```



```

Db 361 AFSVTNNVIRLKPENSVSPLEQALSOHAGHGNLIVAKITTKHPALRLAIQILX 420
Qy 861 RLATVAPMSVYACLGNDAAAIKDAFLTRLOSKIEMRIKVMLEFLVAVETQGLIEFL 940
Db 421 RLATVAPMSVYACLGNDAAAIKDAFLTRLOSKIEMRIKVMLEFLVAVETQGLIEFL 480
Qy 941 LNLVKKDSDGSKESLGMWCLHVLIELIDSOQDRVWCPPLLRRAAFHAIAMORR 1000
Db 481 LNLVKKDSDGSKESLGMWCLHVLIELIDSOQDRVWCPPLLRRAAFHAIAMORR 540
Qy 1001 DSAMVLRTRKPFMNTLSPLEFGLTSPSPSESESLIETCALIMKICILEIYVYKSGSD 1060
Db 541 DSAMVLRTRKPFMNTLSPLEFGLTSPSPSESESLIETCALIMKICILEIYVYKSGSD 600
Qy 1061 QSLKDTLKKFSIEKRPAYWSGYKSLAVHVAETEGSSCTSLLEYQMLVSAMRMLIAT 1120
Db 601 QSLKDTLKKFSIEKRPAYWSGYKSLAVHVAETEGSSCTSLLEYQMLVSAMRMLIAT 660
Qy 1121 HADIMHLTDSVVRQLFDVLDGTAKALLVPASVNCRLRSGMKCTLLILRQMKRELS 1180
Db 661 HADIMHLTDSVVRQLFDVLDGTAKALLVPASVNCRLRSGMKCTLLILRQMKRELS 720
Qy 1181 VDEILGPTEILEGYLOADQOLMEKTKAKYFSAFTVLQMKMKYSDIPQYQSLVLANCE 1240
Db 721 VDEILGPTEILEGYLOADQOLMEKTKAKYFSAFTVLQMKMKYSDIPQYQSLVLANCE 780
Qy 1241 TLQEEVIALFDQTRHSLALGSAITEDSDMETDCSRSHRRDRDQVGVGLHLAECEV 1300
Db 781 TLQEEVIALFDQTRHSLALGSAITEDSDMETDCSRSHRRDRDQVGVGLHLAECEV 840
Qy 1301 DEDGSMVQVTRRLILPTLLTLEVSLRMLKONLHTEHTLHLLTLATQOGATRVAGA 1360
Db 841 DEDGSMVQVTRRLILPTLLTLEVSLRMLKONLHTEHTLHLLTLATQOGATRVAGA 900
Qy 1361 GITOSICPLISVYOLSTNGTAOTPSASRKSIDAPSPGVYLSMSLMEQLKTLRYNPL 1420
Db 901 GITOSICPLISVYOLSTNGTAOTPSASRKSIDAPSPGVYLSMSLMEQLKTLRYNPL 960
Qy 1421 PEALDFVGHQERTLOCLNAVRTVOSLACLEADHVTGFIQLSNFMKEMHFLPOLMD 1480
Db 961 PEALDFVGHQERTLOCLNAVRTVOSLACLEADHVTGFIQLSNFMKEMHFLPOLMD 1020
Qy 1481 IQVNIQYLCQACTSLHSHRMTQHTIQNKXGGLPSAVQORVORPPSASAAPSSSKQPA 1540
Db 1021 IQVNIQYLCQACTSLHSHRMTQHTIQNKXGGLPSAVQORVORPPSASAAPSSSKQPA 1080
Qy 1541 ADTEASBOQALHTVOYGLIKITLTAALRHPTPDVCOILLDOSLDLAEYNFLPALSFT 1600
Db 1081 ADTEASBOQALHTVOYGLIKITLTAALRHPTPDVCOILLDOSLDLAEYNFLPALSFT 1140
Qy 1601 PTFDESVAPSGFTLLATVAVNALMGLDKKEPLTOAVGLSTOAGRTLLKSLMFTME 1660
Db 1141 PTFDESVAPSGFTLLATVAVNALMGLDKKEPLTOAVGLSTOAGRTLLKSLMFTME 1200
Qy 1661 NCFYLLISQAMRYLADPAVHPDKQPMKQELSSSELSTLSSLSRYRRGAPSSPATGVP 1720
Db 1201 NCFYLLISQAMRYLADPAVHPDKQPMKQELSSSELSTLSSLSRYRRGAPSSPATGVP 1260
Qy 1721 SPQKSTSLSKASPSQEPQLIOLVOAFVRRMOR 1753
Db 1261 SPQKSTSLSKASPSQEPQLIOLVOAFVRRMOR 1293

```

## RESULT 6

```

ID 09BS12 PRELIMINARY; PRT; 853 AA.
AC 09BS12;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE NUP186 protein (fragment).
GN Name=NUP186;

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OC NCBI_taxid=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RX MEDLINE=22388257; PubMed=12477932;
RA Struhsberg R.L., Pelngold R.A., Grouse L.H., Derge J.G.,
RA Klammer R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusha K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Prange C.,
RA Brownstein M.J., Ueda T.B., Tothyluk S., Carlini P., Schetz T.B.,
RA Rane S.S., Loguercio N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bock S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalske U., Smalios D.B., Schnerch A., Schein J.B.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RA Struhsberg R.;
RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC005407; AA005407.1; -.
FT NON TER
SQ SEQUENCE 853 AA; 94944 MW; F9EBD8397BE7ED004 CRC64;

```

Query Match 45.4%; Score 4093; DB 2; Length 853;

Best Local Similarity 96.2%; Pred. No. 7.4e-265; Indels 6; Gaps 1;

Matches 810; Conservative 5; Mismatches 21; Indels 6; Gaps 1;

```

Qy 683 VTIAFLRLITLVKQGLSTQSGVPCMFVLEKMLPSYHMKRYNSHGVREQLCLILE 742
Db 1 VTIAFLRLITLVKQGLSTQSGVPCMFVLEKMLPSYHMKRYNSHGVREQLCLILE 60
Qy 743 LIHALINTCHETDLSHSTPSIQFLCISLAYTEAGQVYINIMGIGVDTIDVMAAOPRS 802
Db 61 LIHALINTCHETDLSHSTPSIQFLCISLAYTEAGQVYINIMGIGVDTIDVMAAOPRS 120
Qy 803 DGAEGGGGOLLIKVYKLAFTVNNVIRLKPENSVSPLEQALSOHAGHGNLIVLAKY 862
Db 121 DGAEGGGGOLLIKVYKLAFTVNNVIRLKPENSVSPLEQALSOHAGHGNLIVLAKY 180
Qy 862 IYHKDPAFLRLAIOQLKRLATVAPMSVYACLGNDAAAIKDAFLTRLOSKIEMRIKVM 922
Db 181 IYHKDPAFLRLAIOQLKRLATVAPMSVYACLGNDAAAIKDAFLTRLOSKIEMRIKVM 240
Qy 922 LEFLTVAETQGLIEFLNLEVKDSDGSKESLGMWCLHVLIELIDSOQDRVWCP 982
Db 241 LEFLTVAETQGLIEFLNLEVKDSDGSKESLGMWCLHVLIELIDSOQDRVWCP 300
Qy 982 LIHRAAIAFLMALMDRDSANLVLRTPKPFMENTSPLEFGLTSPSPSESESLIETCAL 1042
Db 301 LIHRAAIAFLMALMDRDSANLVLRTPKPFMENTSPLEFGLTSPSPSESESLIETCAL 360
Qy 1043 IKKICLRIYVYKSGSDLSKDTLKKFSIEKRPAYWSGYKSLAVHVAETEGSSCTSL 1102
Db 361 IKKICLRIYVYKSGSDLSKDTLKKFSIEKRPAYWSGYKSLAVHVAETEGSSCTSL 420
Qy 1103 EYQMLVSAMRMLIATTHADIMHLTDSVVRQLFDVLDGTAKALLVPASVNCRLRSGM 1162
Db 421 EYQMLVSAMRMLIATTHADIMHLTDSVVRQLFDVLDGTAKALLVPASVNCRLRSGM 480

```

```

QY 1163 KCTLLTLTLTROMKRELGSVDEILGLTLELGVLOADQOLMEKTKAVSAFTVLOMKE 1222
DB 481 KCTLLTLTLTROMKRELGSVDEILGLTLELGVLOADQOLMEKTKAVSAFTVLOMKE 540
QY 1223 MKVSDIPQYSQVLYNVCETLOEVIYALFDOTRSHALGSATBEDOSMETDDCSRRRDQ 1282
DB 541 MKVSDIPQYSQVLYNVCETLOEVIYALFDOTRSHALGSATBEDOSMETDDCSRRRDQ 600
QY 1283 RDCVVCVGLALAEELCEVDPDGSOLVOTRRLPLPLTLTLTLEVSLMKONLHFTETATLH 1342
DB 601 RDCVVCVGLALAEELCEVDPDGSOLVOTRRLPLPLTLTLTLEVSLMKONLHFTETATLH 660
QY 1343 LLLTLARTQOGATAVAGAGITOSICPLSLVYOSTNGTGTOTPSAKRSKLDAPSWPGVR 1402
DB 661 LLLTLARTQOGATAVAGAGITOSICPLSLVYOSTNGTGTOTPSAKRSKLDAPSWPGVR 720
QY 1403 LSWSLMEQLLKTIRYNFLPEALDPVGVHQBRTLOCLNAVTVOSLACLREADHTVGFILQ 1462
DB 721 LSWSLMEQLLKTIRYNFLPEALDPVGVHQBRTLOCLNAVTVOSLACLREADHTVGFILQ 780
QY 1463 LSWPMKEWHPLPOLMRDIOVNLGYLCOACTSLHSKMKLQHTLONKNGDLPSSAVARV 1522
DB 781 LSWPMKEWHPLPOLMRDIOVNLGYLCOACTSLHSKMKLQHTLONKNGDLPSSAVARV 834
QY 1523 QR 1524
DB 835 ER 836

RESULT 7
08C7A1 PRELIMINARY; PRT; 798 AA.
AC 08C7A1;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched
DE library, clone:DJ30017D22 product:hypothetical protein, full insert
DE sequence.
GN Name=BC025526;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=21085660; PubMed=11217651;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RA The FANTOM Consortium;
RT The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okaaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to

```

```

RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=20530913; PubMed=11076661;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama U., Nishi K., Katsunai T., Taahito H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara B., Watabiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai U.,
RA Okaaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RA Adachi U., Aizawa K., Akiyama T., Arahawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kaikawa T.,
RA Katoh H., Kawai U., Kojima Y., Kondo S., Kono H., Konda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Mureta M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okaaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK052272; BAC14911.1;
DR MGD; MGI:2446190; BC025526.
KW Hypothetical protein.
SQ SEQUENCE 798 AA; 90029 MW; BA33JAB589BAFAC CRC64;
Query Match 43.7% Score 3939; DB 2; Length 798;
Beet Local Similarity 95.4%; Pred. No. 1.4e-254;
Matches 751; Conservative 21; Mismatches 15; Indels 0; Gaps 0;
QY 15 RSSRELWTLILGRSALRELISOIEALNQRWRLLEGLSYKPPSPSSAEKYANRQVAVSP 74
DB 11 RSSRELWTLILGRSALRELINQIBALNKYQWRLLEGLSYKPPSPSSAEKYANRQVAVSP 70
QY 75 LKELGLRISKFLGUDDEBSVOLLOCTLOEDYRGTBDYVYTVLODRGSOALLTKIADYY 134
DB 71 LKELGLRISKFLGUDDEBSVOLLOCTLOEDYRGTBDYVYTVLODRGSOALLTKIADYY 130
QY 135 EERTCIACVLAHLTYFQDERHPYRVEYADCVDLKEKELVSKNRQOFEELVKTAPVET 194
DB 131 EERTCIACVLAHLTYFQDERHPYRVEYADCVDLKEKELVSKNRQOFEELVKTAPVET 190
QY 195 HGNLMTREVSRWFOCIQREOSMLLEIFLYAYFEMAPSDLVLTAKFKEGSPGSSRQTN 254
DB 191 HGNLMTREVSRWFOCIQREOSMLLEIFLYAYFEMAPSDLVLTAKFKEGSPGSSRQTN 250
QY 255 RHLVDEMDPPVDKIGFSAIILYBGNDISLHAKCALDDBRELHOPADGLICQDMDCM 314
DB 251 RHLVDEMDPPVDKIGFSAIILYBGNDISLHAKCALDDBRELHOPADGLICQDMDCM 310
QY 315 LTPGDIPIHAPVLAAMLILRHTLNPEBTSVVRKIGTAIOANFOYVTRLLQSLASGN 374
DB 311 LTPGDIPIHAPVLAAMLILRHTLNPEBTSVVRKIGTAIOANFOYVTRLLQSLASGN 370
QY 375 DCTSTACMCYGLSLFSLTSLATLGNODIIDTACEVLADSLPELFWGTEPTSGLG 434
DB 371 DCTSTACMCYGLSLFSLTSLATLGNODIIDTACEVLADSLPELFWGTEPTSGLG 430
QY 435 IILDSVCGMPHILSPILQILRALVSGSKTAKKYISPLDKMSFNLKXKHPHDVISHD 494
DB 431 IILDSVCGMPHILSPILQILRALVSGSKTAKKYISPLDKMSFNLKXKHPHDVISHD 490
QY 495 GTLWRBQPTLLVPLVGGQTNLRIPQGVGVMDDBRAVLYVWMEYSWSWTLFTEIEMLL 554

```







Db 972 MEBFLAISNDPAK--LASPLLSR-IMSLFHALMKNMNOALVKGLLOKXNFMSSLCNPLF 1028  
 Qy 1023 GTLSPPSTSEBSILETCALMKITCLEIYY 1054  
 Db 1029 S-----SNNRSYSOLFNIIGIEHFRV 1053

RESULT 11  
 ID 089083 PRELIMINARY; PRT; 190 AA.  
 AC 089083;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 GN Name=BC025526;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=BALB/c; TISSUE=Spleen;  
 RX MEDLINE=99012997; PubMed=9798653;  
 RA Chu C.C.; Paul W.E.;  
 RT "Expressed genes in interleukin-4 treated B cells identified by cDNA  
 representative reference analysis.";  
 RL Mol. Immunol. 35:487-502(1998).  
 DR EMBL; U89435; AAC36539.1; -  
 DR MGD; MGI:2446190; BC025526.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 FT NON\_TER 190  
 SQ SEQUENCE 190 AA; 21250 MW; EA00A3979BBD2B CRC64;

Query March 9.2%; Score 829; DB 2; Length 190;  
 Best Local Similarity 86.3%; Pred. No. 2.7e-47;  
 Matches 164; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

Qy 1251 DQTHSLALGATGDKSDMETDDCSRRHQRDQDVGVGLHLAKELCEVDEGDSWLOV 1310  
 Db 1 DQTHRLVSDSAEDKDMETDGCRRHQRDQDVGVGLHLAKELCEVDEGDSWLOV 60

Qy 1311 TRRLPILFTLTLEVSILMKONLHFTBATLHLTLTARTQGAFAVAGAGTOSICPL 1370  
 Db 61 TRRLPILFTLTLEVSILMKONLHFTBATLHLTLTARTQGAFAVAGAGTOSICPL 120

Qy 1371 LSVYQLSTNGTAQPSASRKSIDAPSWFQVRLSMLEOLKTLRYNPLPALDFGVH 1430  
 Db 121 LSVYQLSTNGTAQPSASRKSIDAPSWFQVRLSMLEOLKTLRYNPLPALDFGVH 180

Qy 1431 QERTLOCLNA 1440  
 Db 181 QNRTWQPFNA 190

RESULT 12  
 ID 09V6F6 PRELIMINARY; PRT; 1822 AA.  
 AC 09V6F6;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE CG8771-PA.  
 GN ORFNames=CG8771;  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratides P.G., Scherer S.E., Li P.W., Hoekstra R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazer V.H., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck U., Brokstein P., Brotlier F.,  
 RA Burlingame K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo J.B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferrar C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jatalat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,  
 RA Palazzolo M., Peltzman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shu B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtk R., Tector C., Turner R., Venter A.H., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wasserman D.A., Weinstock G.M., Weisbach J., Ye J.,  
 RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers R.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).

[2]  
 RC SEQUENCE FROM N.A.  
 RP MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren B.J.,  
 RA Svrtskas R., Taber P.E., Wan K., Stapleton M., Sutton G., Venter C.,  
 RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
 melanogaster euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).

[3]  
 RC SEQUENCE FROM N.A.  
 RP MEDLINE=22426070; PubMed=12537573;  
 RA Krommiller B., Bergman C.M., Krommiller B., Carlson J., Svrtskas R.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the Drosophila melanogaster euchromatic  
 a genomic perspective.";  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).

[4]  
 RC SEQUENCE FROM N.A.  
 RP MEDLINE=22426069; PubMed=12537572;  
 RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Berman B.P.,  
 RA Beaton C.B., Celniker S.E., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 systematic review.";  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).

[5]

```

RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN (6)
RE SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003821; AAF58471.3; -.
DR FlyBase; FBgn0033766; CG8771..
SQ SEQUENCE 1822 AA, 208406 MW, E2C45756B8AD70 CRC64;

Query Match      8.9%; Score 801.5; DB 2; Length 1822;
Best Local Similarity 22.1%; Pred. No. 6.3e-44;
Matches 409; Conservative 319; Mismatches 769; Indels 351; Gaps 69;

QY 2 TRKSITVSLSPCRSSREIMTLIGRSARELSQIAELKMKRRLBGLSYKKPSPSS 61
DB 4 VEKSVIT-----DWKRLMPVSGIHYETTPDUTVARELMNVASLEQAGVLPKPKNASS 56
QY 62 AE-----KYKANKDVASPLKEIGLARIKELGLDEBQSVOLLOCYQEDYGRTRDSVQTVL 116
DB 57 LELGLTLKEKQKQKEL-PTFE---RLQDLIDLESAGCWEILCYLTQETRGASASLITOLI 112
QY 117 QDEROSQALIKIADYYTEERTCILRCVLLHTYFODESHPRVREYACVDKLE-KEIVS 175
DB 113 STEIEMAKIHEDIIRHYSLERWVVLKIVQLIVFHQVPMHPYRERAVAEKITTPLRLAD 172
QY 176 KYRQOFEEIYKTEAPTWETHGNLM--TERQVSRMFWQCIREQSMLEITFLVYAYEEMAP 233
DB 173 SYLDLDESLICVPRPKAKMAECFHSABEIVA-WSEKNAKEITNEVAILLVIAEHLPMGL 231
QY 234 SDLLVLTWFKEQGFGSRQTNHLYDETW---DPFVDRIGYFSALILVEGMDIESIAKCA 290
DB 232 EQIKRIIPAACKQHSFGKQKS---YLDPSQPYHQEIRLSYSGLMVLKCLDPE----- 282
QY 231 LDDRELHQFADGGLICQMDCLMTFGCIPIHAPPLALMAL--LHRTINPETSQV-- 346
DB 283 ---KEEKISDLLEKI-EDLOVDIASMYHRPHEGPLLAMMLLRGTINADADASSILRC 338
QY 347 RKIGGTALQVNFYVLTRLQSLASGNDCTTACMCVYGLSFLTSLLEH-TLGNQO 405
DB 339 ROLGKRAVDLKCFTVOLHIAHSMWADBMLSRIYRTIYNQGYLCLFDDGSGCARYE 398
QY 406 DIIDTACVLAADPSLELFWGTEPTSGLIILDSVCGMPHLLSPILQILRALVSGKSTA 465
DB 399 GIYELLCGLSWFHLAKOFCRSRELT-----HLSKALSTTKA-----GQG 439
QY 466 KKYVFFLDKMSF---YNELYKAKPHDIVSHEDGTLMRQRTKLLYPLGGQTNLRIPOST 521
DB 440 NYVKSQALPILALRYDE-SQHKUREVDITNEFELLASVQPO-----QIDFTIPAGT 491
QY 522 VQGVWLDRAVLYVREYSSSWTLFTCEIEMLL-----HYVSTADVIQHQRKPIIIDLVH 577
DB 492 SCTAIQHSGSCPMFRFPVNYFDALHHEINCLARTETGHLHGFESSERIRNVEAGIRFJE 551
QY 578 KVISTDLSI---ADCLLPITSRIYMLQRLTTVISPPVDVASCNCTIVLAANPAKV 633
DB 552 SAVKLSSQISIGISAMVHP-TEMCVDLHTHTFKSVQYPPYGLSSGLNVTALLPLVDEBI 610
QY 634 WTDLRHTGFLPRV---AHPVSSLQMSISAEGNAGGYGNLMSQPOGEYQVT---IAF 687
DB 611 FSRISNHLILPVPSPGSHYDFGMVYANANGVFESRFLGSIVINDVEKKEERYEFLLSYIGF 670
QY 688 LRLITTLVKGQUGSQGLVPCWVFLKEMLSYHKHRYNSHVGREOIGCLILELHAI 747
DB 671 LRAVSNLKRNR---QIQWEIPLGLIFLLKDVPHLHTHFSQVERENKIYEFILSPICDI 726
QY 748 LNL-----CHETDLSSHSTPSLOFLCICSLAYTEAGQVYINIMIGIVDTIDMMVMAAQ 799
DB 727 LDLFTAKESNCKQREL-----LVKVCVVSILNLENGLILIRFVGVGNNAVQYTMLE 779
QY 800 PRSDAEGQGGQGLIKTVKLAFSVTNNVIRLKP---SNVVSLEQALSQGHAKGNL 855

```

```

DB 780 ---TNMGOQPHGIML--VRLSMRIIMQLRLKEEYVGNSETLSLEALITYQPKORDYL 835
QY 856 ---IAVLAKYIYKHKHPALPRALIOILKRLATAPSPSVYACGNDAAIADAPLTRQSKI 913
DB 836 RIIPTVCSTMSNIFPRMLPILSCRLLKRIALQFNNSLACIMEDQILHTTMQKLPDEL 895
QY 914 EDMRIKMLEFLTYAVETOPGLIELPLN---LEVKQSDSKSEF--SIGMMSCLHAYL 967
DB 896 BSDSIKIMILELVADACIAKQPEVTEAFKRVNVALDKRGRSPFSKOCVPIGR-SITYVNR 954
QY 968 ELIDSGQODRYWCPPLHRAIAIPLALMODRDSAMVLTKPKFMENTISPLFCTLSLP 1027
DB 955 DFLDLQVDPPLTIOALPAKIMTIFHSW--KHNLQMLV----- 991
QY 1028 PSETSPSILETCALIMKICIEIYYVWKSIDSLQDKTLKESJIKRPAYSGYVKSIA 1087
DB 992 -----DILLNIIISIEV--TGNGNNALLDVNKFPEQKFGPMLNAYFNP 1037
QY 1088 VHAETEGSSCTSLLEYQMLVSAMRMLIATTHADIMHLDTSVVRQLFLDLDOTKAL 1147
DB 1038 KVPAYKMLSSBDHPDWCCLQAFKDLIVY-----LKKQPKF 1075
QY 1148 LVPASVNCIALGSKCTLLLI-----LIRQMKR---ELGQVD 1182
DB 1076 VTIPESQ--FKLMQKCLVIVDRSNYLEDKRPPIILAEVYFILLRFGHAYTDSLBEQO 1133
QY 1183 ELIGPLTEILSGVLA--DOOLMEK-----TKAKVSAFTVYQMKEMKSDIPOYSQ 1233
DB 1134 TLMDDLQIMRICALCTEDQVRAKACLAIVTK--THLYTDLIIDSJA-LRPLNS 1189
QY 1234 LVLVNCCITLOEBVIALFDQTRHSIALGSAEDKDSMETDSCRSRRHQRDQVCYVGLTL 1293
DB 1190 VVGICSELEQW-----ENSVLSKSGGLNNSDSKSTYNS-----LITCLNL 1234
QY 1294 AKELCEV-DEQG-DSMLQVTRRLPLPTLTLEBLSLBMKQNL-----HPEBATL- 1341
DB 1235 LKAVATIFRNDGPGW-----DLRP-----VSVELFRLVACVSRTLPFLSKQVLS 1280
QY 1342 -HLTLTLARTQOGATV--AGAGITQISICPLLSVYQC--STNGAQTGPSARSKSLDAPS 1396
DB 1281 VOILDVILVPAKHCVSFELHCDVGRYIMKLLPRLILQSKHETKTTAADABEGWYEQ 1340
QY 1397 WPGYRLSMSIMEOLKTLRYNPLPEALDPYGVHQERTLOCINAVTYQSL--ACLEBAD 1454
DB 1341 WMPVYARIGIEVITIIYEGKCKCFLEDAFOGIAHVPLEDAL--LLSKQSLSEPMAYLIK 1398
QY 1455 HTVGFILQISNPKEMW---HFHLPOLMRDIOVNLGYICQACTSILHSRKLQHTLQKN 1510
DB 1399 AAVNLVASLSTEHKWKQSDSLANIMRAVO--SLLCHT-SLSEFHQKMLKCLL----- 1450
QY 1511 GDGLPSAVAGVQRPSPASAPSSKOPADTEA-SEQALHTVQYGLIKLSKTLAAL 1569
DB 1451 -----AGRSQLEILASTELIYDDELISACNDLTDIIISCVARL 1490
QY 1570 RHFTEDVCOILLDOSLDAEYNFLFALSFTTPTPDS-VAESFGTLATVNV--ALNN- 1624
DB 1491 LRSPDLMEILLCASYESKSKSILLDYVFGAPKLBENBULTFPGVLMVLMVYKALNNQ 1550
QY 1635 -----LGEIDKCK-----EPLQAVGLSTQABSTRILKSL 1655
DB 1551 NHGFSVPPLNSLPVNEHSGDDNDPEVCVGNQTNTPSPSLN-VSISF---GTCPASBIL 1606
QY 1656 -----MPTMENCQFYLLISQARBYLDPVAVHPRDKQRMKQBSSEL 1695
DB 1607 SNMDGCLLALBHLMLVASQALCITRSPULETLMKQIVRDISNEL 1654

RESULT 13
QEDRN9 PRELIMINARY; PRT; 1997 AA.
AC QEDRN9, 01-OCT-2004 (TREMblrel. 28, Created)
DT 01-OCT-2004 (TREMblrel. 28, last sequence update)

```

DT 01-OCT-2004 (TREMELREL. 28, Laet annotation update)  
 DE C70rf14-like.  
 OS Brachydonto reio (Zebrafish) (Danio reio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 CC Cyprinidae; Danio.  
 NCBI TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Amertem A., Nissen R.M., Sun Z., Swindell E.C., Farrington S.,  
 RA Hopkins N.; Essential for Early zebrafish Development.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2004).  
 RL EMBL; AY648720; AAT6038.1;  
 DR SEQUENCE 1997 AA; 223244 MW; 933087C347CEC351 CRC64;  
 SO  
 Query Match 4.1%; Score 372.5; DB 2; Length 1997;  
 Best Local Similarity 19.1%; Pred. No. 3.9e-15;  
 Matches 420; Conservative 296; Mismatches 799; Indels 685; Gaps 98;  
 QY 1 MIRSKITSVLSFCSSRELTITLG---RSALRELSQTEALNCHMRLEGLSYKXP 57  
 DB 1 MATQAVNAGSALWGPPLKEALVEGAIWRQPEVHLLDQLKGRKPYF---LNLKXP 57  
 QY 58 SPSSAEKVKANDVASPLKELGLRISKFLGLDEBOSVOL-----OCYLOEDY---RGR 109  
 DB 58 PKSAEQRKVRK--AS-----TEGALOGGAGARLLPQOLISEAFILSDLPDGL 106  
 QY 110 DSVKTVLDERG-----SQLILIKIADYYEERTCI---LRCVL----- 145  
 DB 107 AALBELLAGEHQHPFGLTRGLVAVL---LYWDGKRCVANSLSRLSIOSRHKFTLIDSLP 164  
 QY 146 ---HLTYPODERHRYREYVADCVDKLEKELVS---KTRQOFEELKTEAPMETHTGNLMT 200  
 DB 165 ELVNLTFETBEL-----MAGLTKQILNIVSDVSTREFRERLCKERG---LGNEXH 213  
 QY 201 EROVRWFOCUREOSMLEIFLYAYAFEMAPSLDLYVTKMFKEGSGRQTNRLVDE 260  
 DB 214 RKEVSDLIKEC---RSLAECLPANAOSPLGKDTLALI-----GHLEMYA 258  
 QY 261 TMDPFDRIGYFSALILVEGMDIESLHKCALDRRELHQ---FAODGLICODMCLMLT 316  
 DB 259 EADGSLDSVNLALVVALYSLDVSFLAQ-GTEDREDLQALPLTEKQVVAVHSRLVAG 317  
 QY 317 FG-DIEH-HAPVLLAMALIRHTLN--PE-----ETSSVVRKGGTRALQANVQYLTR-LL 366  
 DB 318 QGWKPLGLOAVVQVLMALSLRALSQLPQCALVEFEADEALADQLAGVFLPTEGLT 377  
 QY 367 QSLASGMDCTSTACMCVYGLLSFVLISLEHTIGNOD-----IIDTACEVLADPSLPE 422  
 DB 378 GSDGFSQEEFYTRLHSLITDFLA---LMPKVKQLNRADEDARLIHMLQW----- 427  
 QY 423 LFMGTEPFGIILD-----SYCGMP----- 444  
 DB 428 ---GSEPFSSSLKQIDHMLILGEFYSKOPFELELALFEWCPSBSLQHSLSLGSFLGVPL 484  
 QY 445 ---PH-----LSPILLQILRALVSGSKTAKKYSFLDKMSFYNELY 482  
 DB 485 ORPHTTEVVLKSFVROWGDLPLATLYIPYLRMMKGLANGPQCSHVCFSILKT----- 536  
 QY 483 KHKPFDVSHEDGTLMRQTPKLLVPLGGQTLRIPOGTIVGOVMDDRKXIVYRWESYS 542  
 DB 537 NGAPR-----GGRNQ-----AGVSGS-----LVSWHEFFFS 562  
 QY 543 WTLFTCEIEMLLHVSTADVIOHCORVKPIIDLVHKVISTDLSIADCLPISRIYMLLO 602  
 DB 563 LMLYH---ENLRADVPASDSTOY--RHLPRIQITQREL-----EGLTAFQ 603  
 QY 603 RLTIVIS-----PYDVIVASCVC-----LTVLAA--RNP--AK 632  
 DB 604 LITTIITSENARLALCEHGWTPVVMGLGLOCSQVPLKAQVHLVLAFAKSPRIAS 663  
 QY 633 VMTDLRHGFLPFAVHPVSSLSQMTSABGMNAGYGNLMLNSEQPOGEGVITAPRLIT 692

DB 664 LWOSELYTOILQTVKIPQORQAAGIEV-----LNEIESSCEBYPLPRACHIS 713  
 QY 693 TLVKGQL-----GSTQSGVLPCWFWFKEMLPYHKRYNSHGVRBOIGLILIELHAI 747  
 DB 714 TLVESALPVNIGAGLAFAPQDYDLFDLSDVFLAPPTAIRRSARKMEVAEVLVEFKL 773  
 QY 748 LN-----LCHETDLSHTP-----SLQF-----LCICSLAYTEACQTVIN 783  
 DB 774 LREYEPQSDPLPENVEIQSGVPAHMRPGHSLMHLNDSPTSLC-----LN 822  
 QY 784 IMGIGVDITDMWMAOPRSDAEGGQGLIKTVKLAFSVT--NNVIRKPPSNVSPLE 842  
 DB 823 LLEBGAQLD--TYAQPEPKQLESAYLHCLCLELPLQKVTYFMDLIRESQSLIVSPLE 881  
 QY 843 QALSQHGAGNNL--IAYLAKIYH-KNDPALPRLAIOQLKRLAVAPMSV----- 890  
 DB 882 QLLQGSVAKSRKADHTIARIYLYHSSNPDAFQSAKILRIRITRYPNIOARLVGDFTHD 941  
 QY 891 -----YACLGNDAA-----AIRDAFLTRLOSKIEDMRKVMILEFLTVAVE--TOP 934  
 DB 942 QAVSERLMAQFVCELDSEBQEGVTTNDSSEKRAKIRH--ETKIHILMLITSIELTOP 1000  
 QY 935 GLIELFLNLEYDGDGSKFESLGMM---SCLAAVLELIDSGQDRYCPPLHRA--- 987  
 DB 1001 NLGLVILGYEVKKPVSTNLQDPVILGCPSCIAHILSLQGRSDSRSG--PELIRQAQPL 1059  
 QY 988 ---AIFALHMDPRDSANLVARTKPKFRENLTSLPFGTSLP--PSETSPTSI--LET 1039  
 DB 1060 AELCYQVYIOLACACDTSPTMYRTYRTSODF-----LFSHLQHLPVLSSENEALASQ 1112  
 QY 1040 CALIMKICLEIYVYVKSGLDOSLDTLKKSIEKPAWMSGYVSLAHVATEGSSCT 1099  
 DB 1113 MSWLMKTAIRKRYT---SLNRQSHYORLNL-----LDDQPHLITGGE--T 1157  
 QY 1100 SLELYQMLVSARMILLIATTHADIMLTDVVRQ-----LELDVDTKALLLVPAV 1154  
 DB 1158 GMEENRNSVSGFLQGTVSKVRRLLSVLDADIFQABELLQDLFFERTQIEQYI---T 1214  
 QY 1155 NCLR--GSMKCTLLILRLQMKRELSVDEI--LSPLELLEBVLQADQDMKRTAK 1209  
 DB 1215 NCEHNEQGHYVCNKL--LHRVLVAEINALQGMALIGRPLMEBVSNTLQOVVERNRK 1273  
 QY 1210 VF-----SAPITVLQMKEMKVSIDIPQYSQVLVAVCEYQOEVI----- 1247  
 DB 1274 RSLAKRRALDSKWSLVEITILLACPSSELPADQROLITRDLLDHDVLSDAAGELMP 1333  
 QY 1248 ---ALFDQTH----- 1255  
 DB 1334 IVAGAVFTLTHALSGSVSEQGVPEGGSSGSPASINSALHLIRKLDLPILCTGGGF 1393  
 QY 1256 -----SLAGSATEDKDSMET-----DDCSRRHRDQGVCVATGL 1291  
 DB 1394 QELRAHLYGALLYYLQIAQKEPEPFTLQGISMMERLTAPBEDCFKLRQENISIESYGT 1453  
 QY 1292 HIAKELCEVDEG-----DSWLQVYRR-----LPIPLTLTLTXS 1327  
 DB 1454 ALMEVAGDADCGHEIGMLALAVLDRLVSLIDRQOMLVILCNSGYLRVLVSLQDDVS 1513  
 QY 1328 LR-----MKONLHFEATLHLTLARTQOGATAVAGAGITOSI--CLPLSVYQLS 1377  
 DB 1514 LQTLTPQRPPLKPIYTESKALLTRVAKTRAGMELLRCGLVAVQVBC--QVFHML 1569  
 QY 1378 TNGTAQTPSASRKSIDAPSWPQVYRLSMSLWEQLIKTLRYNFLPALDPFGVGHQERTQ-- 1436  
 DB 1570 PQNDALRVFGGD-----PS-----GRIPAPL-----RRYQOI 1597  
 QY 1437 CLANVRVQSLACLEADHTYGFILQLSNPKEM--HEHLPQ--LMRDLQVNLGYLQ--A 1491  
 DB 1598 LIPTRLMQVILITSTTAHQQAQVLT-----QWLIHSDVLOSTILHGDMMGSLQDELS 1652  
 QY 1492 CTSILHSRMLQHYLVONKNGDGLPSAVAO-----RVORPSPAA----- 1529

Db 1653 LITAIISKALPGALB--MGQEINSAALMELQGHIGRFORQSLSLVRLVGTDRARYLKQ 1710  
 QY 1530 ---SAPSSSKOPADPTEBEOQALHTVOYGLKILKISLALRHFPDVOQILIDQSLD 1586  
 Db 1711 IEDTVSPSNTAEREDMEVMOQCANIM-----EYQQTILQSSS 1751  
 QY 1587 LAEYN-LFALSTFTTP---TFDSEVAPFG-TLLATVNAVNLMLGELDKKEPLTQ--- 1637  
 Db 1752 EQGFSLCLPSPSABPADVSIIPSAHVPSLGLVILLNLSGSDPFRYHSHRQSLNKLERY 1811  
 QY 1638 -----AVGLSTQAEGRITLKS-----LIMFTWENC 1662  
 Db 1812 EQLPPEELKELCOGLVSGSGVEKISSVQRNVLAKRRLVOLVNNRAXKLALQSDIITC 1871  
 QY 1663 FYLLISQAMRYL-----RDP--AVHPRDKRMQKQELSELSTLSSLRFRGAPSSPA 1715  
 Db 1872 LFTYHMLERYLYCYPTFDKDSILPGRQSLKLDYFQMLNDVSY-----Q 1922  
 QY 1716 TGVLPSPQO-KSTSLSKASPSQEPFL-IQLVQAFVRHMOR 1753  
 Db 1923 IALCADIAPAGSRGLSVSRVSCQDLEQLQSDMSSSPSEFLQR 1962  
 RESULT 14  
 OBLM71 PRELIMINARY; PRT; 2025 AA.  
 ID 08LM71;  
 AC 01-OCT-2002 (TEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein OSJNAA0034B05.4.  
 GN Name=OSJNAA0034B05.4;  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzae; Oryza.  
 OC NCBI\_Taxid=39947;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Wang R.A., Yu Y., Yang T.J., Nah G., Soderlund C., Chen M., Kim H.-R.,  
 RA Rambo T., Sasaki C., Henry D., Oates R., Simmons J.,  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC12145; AAM46049.1; -  
 DR Gramene; OBLM71; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 2025 AA; 224968 MW; 9F69BD4AA1FE89A0 CRC64;  
 Query Match 4.1%; Score 368.5; DB 2; Length 2025;  
 Best Local Similarity 18.1%; Pred. No. 7.3e-15;  
 Matches 384; Conservative 327; Mismatches 652; Indels 755; Gaps 98;  
 QY 25 LGHSALRELSQTEA---ELNKNHRRLLGSLTYKPPSPSAEKVKANKVNAS----- 73  
 Db 45 LDRAAASBPVPALMERIKAHNAMLGVSVMFVKPDSARGLADSEVVVGGHRLAVKP 104  
 QY 74 ELKELGIRISKFTGLDEBSVOLLOCTLODYGTBSVKTVALDEQSQALILKADY 133  
 Db 105 ELKAAALIRSKCNMLDEVOSYILV-----KRTSNTPTALVAD--TEEFELAVSVQY 154  
 QY 134 YEERTCILRCVAILLTFQDERHPYRVEVADCVDKLEKELVSKYQOFEF-----LYKTE 188  
 Db 155 YLERQCLIKCIRIRIIFYHND-----CSDSIDAVAEASVLRREVEGRLLSTIVDSL 206  
 QY 189 APTWETHGNLMTERRQVRWFVQCLREOSMLEIIFLYAVFEMAPSD--LIVITKPKFE 245  
 Db 207 ASAPSVYKG--AEILTS-WLEBTLIRINLIFDLFF--YDNLSCRNGGLMIMLCIFKO 262  
 QY 246 QGSGSQTRHLYD-ETMDPVDRIQIFSA---LIVEGMDISELHKALD---RREKH 298  
 Db 263 MLSGSYDVGFPAVSVEAKNSF---HYAKQQLPFIQLIDFESLIMVDEVPFSGS 318  
 QY 299 QFADQGLIQDMDCLMTFDGDIPHHAFLVLAAMLLRHTLNPBEERSVVRKIGGTAIQLV 358

Db 319 TFSVVDILEMDVEVSKLPEFAVESGPLILANAVF-----LCLVMSLPSSNTNLP 369  
 QY 359 FOYLTRLLOSASGNDCTTSTACMCVYGLSLFVLTSLBLHTLGNODDIDTACEVLADP 418  
 Db 370 VSGFRILKTFIS-----AVVASY----- 388  
 QY 419 SLPFLMGTSEPTSGLOIINDVCGMPPHLLSPILQD--LRAVSGSTAKVYSPDKMS 476  
 Db 389 ---EISQYEDDS-LGMINILICEVYDGEESLQCMQWDSQFIDGFLR-----FVYLERNN 440  
 QY 477 FVNELY--KHKPRDIVSHEDGTLNRQTKELYPLOGQTRAPRPGTVGVM--LDDRAY 532  
 Db 441 GVTTLAVPRSDTDNNVYHQI--EHSPLSIFGIGTT--IPGSHGILVLEEDVA 495  
 QY 533 LVNWEYSYSWTLFTCEIEMLHVSTADVIQHCQRKPIIDLVHNVISTLSIADCLP 592  
 Db 496 LVNWE-----DLCLALHAKSLAVQASQ--LGIYDHNVRIDIAKIFC--- 537  
 QY 593 ITSRIYMLQRLTT--VISPPVDIVASCNCLTVLAARNPAKYWTLRHTGFLPFAHPV 650  
 Db 538 --TSIFKYVEDFNNAQVMSKTLGMLAEMLSCV-----PYHFNVALDCGF--PIYQ-- 584  
 QY 651 SLSQMTISAGNNAAGYGNULMNSBQPOGEYGTIAFLKILTVLKGQ-----GSRQSQ 705  
 Db 585 ---SGVASSDMLSGALARMLPATSESDSCS-----SLTTVLDPALQVLRKGAAD 635  
 QY 706 GLVPCVMPVLKEMLPYHKKRYNSH-----GVREQIGCLILEL-- 743  
 Db 636 IISFPIIFSVQIYIMVHNMKTKYSRMKLTILKVPVLKSCIOVKSPPSLGGLIMEILL 695  
 QY 744 ---HAIL--NICHTEDL---HSHSTPSLOFLCISLAYTBAGTVINIMIGIYDITD 793  
 Db 696 YDSHSVILHLISMSTQLHSHSGYCHDKOI-----BDIQLVLCGPFIVF 744  
 QY 794 MWAAAPRBDGAGGQGOQLIKTV--KLAPSTNNVILKPSNVVSPLEQLSQRGHA 851  
 Db 745 YMLSNLP-----ERKKSRMLRRVGSKSAFRYGVFVAVKTLA-----FRH 786  
 QY 852 GNNLIVAVLAKY-----IYKHDPALPRLAIQLKRLATVAPMSVYACIAND--- 897  
 Db 787 SCNIIIVLSBFSQPSLEMKTCIIFH-----LOLAPIKLS--QSHYACSNSBDNN 834  
 QY 898 -----AA-- 911  
 Db 835 RTSNKEPMDNDNDIPLKAIQVAAARVPSMLCFPAVYKAQOLMENAVFVVGSSIMRLQTS 894  
 QY 912 -----KIEDMRKIMILFELYAVETOP-----GLIEPLNIEVDGS----- 949  
 Db 895 ISCLIDEVNVKVEVVAIFMLISAAARYQAQRADVALVWLRVHLPGSQSTEAQADSDNSAH 954  
 QY 950 -DGSKEPSLGMMSCLHAVLELI-----DSQQODRYWCPLLRRAAIAFHALMODRDS 1002  
 Db 955 BOSSKTFVLPNPGSNRFLVEQILIGYGRSTELMDR--SPSILSGVLDLKLALMESAOQF 1011  
 QY 1003 AMLV--LRTKPKEMNLTSPLEFTL--SPPSST--SEPSILBTCL--IMKICIEIY 1053  
 Db 1012 IYILEKLSRRTFMENTLSCIRAPASYPIDSVETVDKKSIAKCYCGITFELMSYELFL 1071  
 QY 1054 VVK-----GSLDQ-----SLKDTLKRSIEKRPAYMS 1080  
 Db 1072 QGKLTLETQTSDDPAPVSGKQKPSVAPCPSPDVLKMPDSTWEDIDVNHLS-----N 1124  
 QY 1081 GYKSLAVNAET-----EGSSCTSLIEYQMLVSA-----WRML 1115  
 Db 1125 GYQNDI-LHRAKVASCICIIIRLLTKLSSGDTSLSPSLVKIQLISSKLTMCAGATTL 1183  
 QY 1116 I-----IATY-----HADIMHLDTSVVRQLFLDVVDGFKALLLVASVNCIPLGS 1161  
 Db 1184 CGSPRIASRGNOVHWHRRHGRDAOSTDA-----NAPRGQGVVTTT---CI----- 1227  
 QY 1162 MKCTLLIILIRQKRELSGVDE-----IGPITEL----- 1192  
 Db 1228 --CSSTVTICPSWILQWLFIMDEBNPLADIWLKTLHIALEGRGPIITGPFQELLCPLEF 1285



```

QY 1193 -----EGVQAD-----QQLMEK-----1205
DB 1286 KVFENHPSBQLOKSPPAANGVSLFDVPHIRDELGLMNHNSDMKTYKFAERKLDIMHKA 1345
QY 1206 -----TKAVSFATVLM-----KEMKVSDF-----QVSQVLANVCETLOEVB 1246
DB 1346 NLMKQVADKCALMSFITFLVSCGTSTSKYKFGPBGGSISTTQOSAVRCACKSLQSV 1405
QY 1247 IALFDQTHS-----LALGATEDDS-----METDDCSRS 1277
DB 1406 DSLPEVNSGVLPPLSGVVELLTITRLILDHAKQSKSRHLYPVITLMTSGASYS 1465
QY 1278 -----RHRODRG-----VCVGLHAKEL 1297
DB 1466 FLFNMPSSPALQPVKSLVLLSLFEFTYKKVDKDSDEVNIFGELISLSMLPYL 1525
QY 1288 CEVEDGD-----SMLOVTRRLPILPILTLTEVSLRMKN--LHFEATL 1341
DB 1526 CKLASREVPDLAISMIDILKGFPSNVWVPILOHFP--RLQVILQKQSGALLCTQVTL 1584
QY 1342 HLLTLTAQTQAGATAVAGITQISICLPLLSYVQSTNGTAQTSPASRSKLDAPSWPGY 1401
DB 1585 NFLTMTGRKQKAKILQSANIFAFIKV--LLS--QNSLDDSCLRNSLSTQTKVKTW---- 1637
QY 1402 RLSSLSMEQLKTL-----RYNFPBALDF-----VGVHQRITLQCLNA 1440
DB 1638 GLGLAIVSILNKMDDISRNNSVNSTISFLSGQVPLMSSYLSAOSVNHQSKRTLLQK 1697
QY 1441 VRTVQSLACLEADHTVGFILQSLNPKMKHFLPOLKMDIQVNLGYLQACVSLHSRK 1500
DB 1698 SQT--SLSLSLTEN--LTLTLCTLAKHFPDRTGKGVSESL-----RE 1738
QY 1501 MLOHYL-----QNKXGD-----GLPSAVAQRY--QRP-----SAAS 1530
DB 1739 ITHILAFISRGSEKRGDSPPNNLSFGCPPIKEKMLBEPPLISKYGMWFRPAFSCUL 1798
QY 1531 AAPSSSKOP-----AADTASRQO-----ALHTVOGLKIL--SKTLAL 1569
DB 1799 STPSVSGPPNAGLSLIVRDKNPDSMKQRTETEMLAQVLYAIALIMKFLCSQKEXAV 1858
QY 1570 R-----HF-----TPDVQILLDQSLDAENFLPALSFPTPTDSEVAPSGF-- 1613
DB 1859 RRAEELFLDLAHPPELPPMDILHGLQD-----VVSIVTEVLEANNVSTALNTB 1907
QY 1614 -----LLATVVALNM 1624
DB 1908 TERVQCLLVILETSLYM 1925

```

RESULT 15  
N205 HUMAN STANDARD; PRT; 2012 AA.

```

ID _N205 HUMAN STANDARD; PRT; 2012 AA.
AC 092621, Q86YCI, 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Nuclear pore complex protein Nup205 (Nucleoporin Nup205) (205 kDa
  nucleoporin)
GN Name=NUP205; Synonym=C7orf14, KIAA0225;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Bone marrow;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarayashi Y.,
RA Ohara O., Tanaka A., Korani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
  The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
  analysis of cDNA clones from cell line KG-1 and brain."

```

```

RL DNA Ref. 3:321-329 (1996).
RN [2]
RP SEQUENCE OF 1181-2012 FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Struhsberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rudin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Garinchi P., Prange C.,
RA Raha S.S., Loguella N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Boeck S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalon D.K., Wuzny D.N., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton B., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywicki M.T., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
  and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE OF 85-92; 572-581; 1452-1461 AND 1494-1502, AND
  CHARACTERIZATION.
RX PubMed=9348540;
RA Grandi P., Dang T., Pane N., Shevchenko A., Mann M., Forbes D.,
RA Hurt E.;
RT "Nup93, a vertebrate homologue of yeast Nup96, forms a complex with a
  novel 205-kDa protein and is required for correct nuclear pore
  assembly."
RL Mol. Biol. Cell 8:2017-2038 (1997).
CC -1- SUBUNIT: Interacts with NUP93.
CC -1- SUBCELLULAR LOCATION: Nuclear pore complex (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation-
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D86978; BA013214.1; ALT INIT.
DR EMBL; BC044255; AA044255.1; -.
DR Genew; HGNC:18658; NUP205.
KW Direct protein sequencing; Nuclear protein; Transport.
SQ SEQUENCE 2012 AA; 227918 MW; FF92PC840B723F16 CRC64;

```

Query Match 3.2%; Score 292.5; DB 1; Length 2012;  
Best Local Similarity 18.9%; Pred. No. 8.9e-10;  
Matches 416; Conservative 303; Mismatches 743; Indels 741; Gaps 104;

```

QY 41 NKRMRLLLEG-----LSYK--PPSSSAEKY--KANKDVASPUKELGLR- 81
DB 26 NALWRRQPAVAHLIDILKIKKHPDIFSLKPNPKVVOQHEKVOKASTEGVALQCGGTRL 85
QY 82 -----ISKPLGLDBEQSVQLQC--YLQSDYRG--TQDSYKTVQDERSQALL 127
DB 86 LPEQULKEAFIISDFDIEGLAVALLELAGEHQPHFELTGLVAVL- 134
QY 128 KIADYYEERTGLCVLHL-----TYFODERPYRYEYADCV 167
DB 135 -----YWGSKRCIASIKALIOSRGKTTTSLSPASMTTRFTDELMEQGLTYKVL- 188
QY 168 KLEKELVSK--YRQFEELKYTEAPTWETHGNLMTEROVSRWFOVQCLRSQMLLEIFLY 225
DB 189 -----LVGQIDVNNBEBKQRGKSGSKG-----RKEVSDILKEC--RQSLASBLFAW 235
QY 226 YAFBMAPSDLLVLTMTFEGQFGSRQTRHLVDETMDP--FVDRIGYSALIVGMDI 283

```

Db 236 ACOSPLGKEDTLLI-----GHLERVTVEANGSLDVAIVLALMALLYCFDI 281  
 Qy 284 -----BSLHICALDDRE-----LHQAODGLICODMCMCLTEBDIDH-HAPVL 327  
 Db 282 SFEOSTEERDDMIHQLPILTERQYIATHSRLQDSQLMK-----LGLQATVR 330  
 Qy 328 LAMALLRHTLN--PE-----ETSSVVRKIGGTAIQLNVFOYLTRLLOSILASGENDCTST 380  
 Db 331 LAMALLRGTISQPLPDYALAEFTDEADEMELAIADNVFLF---LMSVAVSEYFOEER 387  
 Qy 381 ACMCVYGLLS--FVLNLSLEHTLGNOD-----IIDTACEVLADPSLP----- 421  
 Db 388 YLRVAVNLITDPLALPMKVQGLRNRADEBAPMIMSMQWGNPPISLRDEHMLLIG 447  
 Qy 422 -----ELFNGTEPT-----SGG-----IILDSQGNF-----P 445  
 Db 448 ELVKONPFHLEALEWMCPTPLQPTLIMGSYLGVAHQRPORQVILSKFRQMGDLBP 507  
 Qy 446 HLLSPLLOLRALVSGSTAKKYVSFLDKMSFYNELYKHKPHDVISHEDOTLMRRQTPKL 505  
 Db 508 TYIPLKRLQGLANPQCHAYCFSL-----534  
 Qy 506 LYPLOGQTNRIPOGTVOVMDLDRAYLVWMEYSYSWTLFTCEIEMLHVSTADVIGH 565  
 Db 535 --KXNGSHVENIQAAGSP-----VSWEHFPHSIMLXH---EHLKDLPSADSVQY 581  
 Qy 566 CQWVPRIDLVKRVISTDLSIADCLPIRSRIYMLQ--RLTVIAP--PUDVIASCYN 620  
 Db 582 --RHLPFRGITOK--BODGLIA--FLQITSTIITWSENARLALCEHPQMTVVVILGLQ 635  
 Qy 621 C-----LTVLAA--RNP--AKWTDLRHTGFLPFVAPVBSLSOMISAEAGMAG 665  
 Db 636 CQIPPLKXELKTLAARPKSPBIAASIMOSLEYTQILOVTRIPSRQALGIEVE----- 690  
 Qy 666 GYNLIMNSEQPOGEVGTIAFLRLITTVKQGLSTQSGI-----VPCWVFLKEMLP 720  
 Db 691 -----LNEIESRCEBYPLTRAFQOLISTVSESFPNGLAGLRPPGFDPYLGFLRDSVFL 745  
 Qy 721 SYHKRRYNSHGVREOJGCLILHAIIMLCHETDHSHTPEL-----QFLCICS--L 772  
 Db 746 RRTTRAPRAAEKWEVAEVLLEVFKLLR-----DIEPQLEDVDFVQFELQDEEII 796  
 Qy 772 AYTBAG-----QTVINIMGVDTIIDWMAAOPRSDAEGGQGLIKTV 818  
 Db 797 AKRPGFSLMTHLINESPYLELALSLBESVKOLD-TYAPFPRKGLERKAVGHCLALML 855  
 Qy 819 KL-AFSTNNVTLRKPPSNVVSPELOAL--SOGHAGNNLJAVLAKYIYHKHDPALPRL 874  
 Db 856 TLQKENLPMDDLRESQALIVCFLEQLQGINPRTKADNVNMI-ARYLYHGN--TNPEL 912  
 Qy 875 AIQLKRLATVAPMS-----VYACIGNAAA--IRDAVLTL 909  
 Db 913 AFSAKILCICISCSNIOIKLVGDFTHDOSISQKMAGFVEICDCEADABEPYLBEGSBL 972  
 Qy 910 OSKIEDMR--IKWILEPFLTVAETOPGLIEF-LNLBYWDSGSGKESLGMW-----SC 962  
 Db 973 EKKLVAIIRERTIHIIMLITSLDECNPPIALYLLPBLKPYSTNLDPGLGCPRTIC 1032  
 Qy 963 LHAVELIDSOODRWYCPPLHRAAIAFLHALMO-----DRDSAMVLRTKPKF 1013  
 Db 1033 LHAIMINILEKTEGR--TGPAVAVRESPOLAELCYOYIYOLCACSDFSGPTMRILRTSQPF 1090  
 Qy 1014 WENTSLPFLGTL--PSETS-EPSTLETCALIMKIIICLEIYYVVGSLDOSIKOT---- 1066  
 Db 1091 -----LFSQLOLYPFSNKEYEISIMLQMSWMLKASIELRVT--SLNRQSSHQRL 1140  
 Qy 1067 -----LKFS-----LEKRPAYMSG----- 1082  
 Db 1141 HILLDMVPKPYSDGEGIEDENRSVSGFHPDTATKVRKIIIMLSDIDPSQELPEPQ 1200  
 Qy 1083 -----YKSL-AVHVAETEG-----SCT 1099

Db 1201 LDFPRAQIEQYIANCEHNLNGQTVANVKELHRYLVAEVNALQGMALIGORPLMEBIS 1260  
 Qy 1100 SILEY-----QVLVSARML--LIIATTHADIMELTD-SVVRQPLFVY-- 1140  
 Db 1261 TVLOQVGNKLLQCLHAGRHLESNRQVLVEIILTRCPDOLLQABERQOLIINDIODVHD 1320  
 Qy 1141 --LDGKALLVPAVNCILRSGMKCTLLIILRQMKRELGSVDIEILG----- 1187  
 Db 1321 KLIDBAQELMPVAGA--VFTLTRHLSQNLVTGKQ-----TSVLGPAABHAYAMLDS 1373  
 Qy 1188 -----LEIIEGVLQADQOLMEKTKAPVSAFTVLOMK 1221  
 Db 1374 CFTSPREBENPLVGFASIGDSSLYILKKLDPIILKTGGG--PQRYVTHYGSLLYVLQTA 1432  
 Qy 1222 E-----MKVSDIPOYSQVY--NVCEITLOREVALPQPTNHSALGSATED--KDSM 1269  
 Db 1433 QRPDEBDTLEAKKTKMWERLTPEDEVPFSLQRENTALIE-----SYGALMEVCRDNC 1486  
 Qy 1270 ETDGCSRSRHRDQDQGVCLGHLAKELCEVDDEDSMLQVTRRLPILPILLTLEVSIR 1329  
 Db 1487 DGHETR-----MLALALDRIVSVQK--QQMLLYSNSGYLVKIVDSLVBDR 1534  
 Qy 1330 MKON-----LHFTBATLHLILTLARTQGAATAVAGATQISICLPILSVYQIST 1378  
 Db 1535 TLQSLTPQPPILKALTYTESMAFLTRVAKIQGALBELRSGVI--VELAQCOYVDMKP 1592  
 Qy 1379 NCTAOTPSASR--KSLDAPSGPVYRLSMSLWQLKTRRYNLPREALDPVG-----VHQ 1431  
 Db 1593 ETDPOSGMGMRPWFIFPVDRIYQIILPALQLCOVITSSMAQHLQAAQVLOPLISH 1652  
 Qy 1432 ERTLOCLNAVRTVQSLACLREADHTVGFILQLSNFKEMHFIPLQMR--DIQVNLGYLC 1489  
 Db 1653 SDTIGALRCQDV--SAGSLQELALTLGITSKAA-----LPGILSELVDVNGSLM 1702  
 Qy 1490 Q-----ACTSLH--SRKMLQHYLQNKQGDGLPSAVAQVORPPAASAAPS 1535  
 Db 1703 ELQGHIGRQRCIGLRSFGSDRLRQFKQDDVNEG-----DRV 1743  
 Qy 1536 SKOPADPEASRQALHTVQYGLKILSKTLAALHFPVQCOILLDOSLAEVYFLA 1595  
 Db 1744 SKKDEIEL-AMQICANMEY-----COSLMDS--SPITQJA 1778  
 Qy 1596 ISFTTPTDSEV-----APFQTLATVNVALN-----MLG 1626  
 Db 1779 VCLFPPLSSETVNRNRPQDQAPVVPYWRPLGDIITYLKQNSMDPFSYDSHRQVS 1838  
 Qy 1627 EL-----DKKEPLQAV-----GLSTQAB-----GTRTKSLMFTM 1659  
 Db 1839 KLQNVBOLEPPDEIKE-LQOSVAPAGVDKISTQKVLARRLVKVINNRKLLSLSCSFII 1897  
 Qy 1660 ENCFYLLISQNRVYLRDAVH--PRDKQRMQELSELSTLSSLSRYRFRGAPSSPATG 1717  
 Db 1898 ETCLEPIL--WRHLEYTLHCPPTDSQ-----DSLPAKSTLFFSRRLQDSFAS- 1942  
 Qy 1718 VLPSPQKSTSL--SKASPEQSEPLIQ-----VOAFVHMR 1753  
 Db 1943 -----ETNIDPRSGLAIVSQHLDQLOADAIAPFBSLQK 1977  
 RESULT 16  
 08666  
 ID 08666 PRELIMINARY, PRT: 1139 AA.  
 AC 08666;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Hypothetical protein OSJNBa0034B05.27;  
 GN ORFNames=OSJNBa0034B05.27;  
 OS Oryza sativa (Japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 NCBI\_TaxId=39947;



```

RN [1]
RP SEQUENCE FROM N.A.
RA Ming R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA Saeki C., Henry D., Oates R., Simmons J.,
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA The Rice Chromosome 10 Sequencing Consortium;
RT "In-depth view of structure, activity, and evolution of rice
RL chromosome 10."
RL Science 300:1566-1569 (2003).
RN [3]
RP SEQUENCE FROM N.A.
RA Buell C.R., Ming R.A., McCombie W.R., Messing J., Yuan Q.,
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC097446; AAM14698.1; -
DR EMBL; AB017058; AAP52188.1; -
DR Gramene; O86666; -
KW Hypothetical protein.
SQ SEQUENCE 1139 AA; 127147 MM; 7F50B31552B599FD CRC64;

Query Match 3.2%; Score 292; DB 2; Length 1139;
Beet Local Similarity 18.7%; Pred. No. 4e-10;
Matches 226; Conservative 220; Mismatches 414; Indels 346; Gaps 56;

OY 25 LGRLALREISOIEA--ELANKMRRLLEGLSTYKPPSPSSAEKVKANKQVAA----- 73
DB 45 LDRALAAPSPVDALAEIRIKAHAMLRGVSVMFKPSDASRGALDSRYVVGSEHRLAVKP 104
OY 74 PLKELGLRISKEFLGDEEQSVOLLOCYLEDYRGTRGDSVKTVLODEROSALILKLADYX 133
DB 105 ELKAAALRLSKCMNDEVOISYLV-----KRTSEKPTALVAD--TEEFRLAVSVQY 154
OY 134 YEEERTCIARCVALHLLTYFQDERHPYRVEXADCVKLEKELVSKYRQOFEE----LYKTE 188
DB 155 YLERQCLAKCIRIRIFVHAND-----CSDSIDAVREASVLYVREVEQRLLSIYRDEL 206
OY 189 APTWETHGNTLTEROVSRWFVQCLAREQSMLEIFLYIYAFEMAPSD---LVLVTMFEK 245
DB 207 ASAFVKGK--ALRTIS-WLEBTLIRINIFILFLFF-YDMISRCNGGLMIMLCISIFPD 262
OY 246 QGFGSRQOTRHLVD-ETMDPFDVRIQYFSA--LIVEGMDIESLHKALDD--RREHL 298
DB 263 MLSGSYDVGKFAVSEAKNSF---HYAKQQLFILQTLDESLKMRVDEVPSSGYS 318
OY 299 QPAQDGLICQDMDCLMTLFGDIPHNAPVLLAMWLLNHTLNPEBTSVVKIGGTALQLVN 358
DB 319 TFSVVDILEMDVSEKLPFAVAVESGPLILAMAVF-----LGLVMSLPGSNNTLGP 369
OY 359 FOYLTRELQSLASGNDCTTSTACMCVGLSLFVLTSLDELHTLGNQODIITPACVLA DP 418
DB 370 VSGFRGLIRTFIS-----AFVASY----- 388
OY 419 SLPELFMGWEPSTSGIILDSVCGMPPHLSPLDQ--LRALVSGSTAKKVVSEFLDKMS 476
DB 389 ---ELSYQTEDESS-LGMITNLICEVVDGSESLCMQWDXSFIDGIR---FNHYERKN 440
OY 477 FYNELY--GKHPDVVSHEDGLTMRQTPKLYPLGCGTNLRIPQTVGQVM--LDDRAY 532
DB 441 GVTTLVAVRSDTDVNNVHDOQ--ELHSPISIFGIGCTT--IPGSHGVIKVLLEDDVA 495
OY 533 LVRMWSEYSSWTLFTCEIEMLLHVSTADVIOHQCRVKPDIIVHKVISTDLSADCLIP 592
DB 496 LVKRWB-----DLCLALHMDKSLAVQASQN---LGVIDGVRVIDIAKIPC-- 537
OY 533 ITRIRYMLQRLTT--VISPVDVIVASCNCLTVLAARPAKWTMDLRTGFPFAHV 650
DB 538 ---TSIFKYVEDFNNAVCWSKTIGMLAEMLSCV---PYHFNVALDCGF--FTQ-- 584
OY 651 SLSLSMTISAEGNAGGYGNILMNSQPOGEYGVTTIAFLRLITFLVKGOL-----GSTQSG 705
DB 585 ---SGVAASDMLSLGSLARMTLPATSDSDSCS-----SLTTTVADPAIQVAKKGAAD 635

```

```

OY 706 GLVPCVMFLKEMLPSSYKMKRYNSH-----GVNEIGCLILET-- 743
DB 636 IISSTIIFSVQYIMNMHMKYKYSRMKTLTKVFDVYKSCIOVQSFSSKLGIIWEILL 695
OY 744 ----IHAIL--NLCHETDL---HSHTPSLQFLCISLAYEAGQTVINIGIGVDITD 793
DB 696 YDSSISHSVLHILMSLTQLSHSGSYCHDLDI-----EDIQVLVCCGPDIVF 744
OY 794 MYMAAPPSDGAEGGCGOQLIKTV--KLASFVNMVIRLKPNSVNSVLEQALQHGHN 851
DB 745 YMLSNLP-----EREKSKRMILRRVGSKAFRYGFVVLVKIA-----FRH 786
OY 852 GNNLJAVLAKY-----IYKHPDPALEPRLAIQLKRLATVAPMSVYACLDND-- 897
DB 787 SCNIIDVLSERQPSIEKRTCLIFH-----LQLAPFKILS---QSHVACSSEDN 834
OY 898 -----AAA-----IRDAFLRLQSG- 911
DB 835 RTSNKPWPTDNTDIFLAKIQVAAAVFSLCETAYAKAQPOLMENNVPVNGSEIWRLOT 894
OY 912 -----KIEDNRKIMLEFLTVAVETDP---GLIELFLNLEVKDGS----- 949
DB 895 ISCIDDEVQXNEVVAIFMLLSAARYQADVALVWLRLVPLPESQSTRAQADSDNSAH 954
OY 950 -DGSKEPSLGMMSCLAHVLIEL-----DSQOQDRVWCPPLHRAAIAFLALMODRDS 1002
DB 955 EOSSKTFYALNPGSGNPRFLVEQILGYIGSTELMDR---SPSILSGVLDLKLKMBGAP 1011
OY 1003 AMLV--LRTKPKFMENTLSPFLGTU--SPPSET---SEPSILFETCAL--IMKICILEIY 1053
DB 1012 IYLERKLSRFFWNLSOCICIAAFASYPIDSEVTEDEKSLRYCCLGTIFIMSEYLP- 1070
OY 1054 VYKGSU 1059
DB 1071 -LQGRU 1075

RESULT 17
O6PDG0 PRELIMINARY; PRT; 2003 AA.
ID Q6PDG0
AC Q6PDG0;
DT 05-JUL-2004 (TRENBLREL. 27, Created)
DT 05-JUL-2004 (TRENBLREL. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLREL. 27, Last annotation update)
DE Nup205 protein (Fragment).
GN Name=Nup205;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RC MEDLINE=2238257; PubMed=12477932;
RA Straubeberg R.L., Pelngold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Scheffer C.M., Schuler G.D.,
RA Altchul S.F., Zeeberg B., Burow K.H., Schaefer C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carinci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.U., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton B., Kerteman M., Madan A., Shcherchenko Y., Bouffard G.G.,
RA Whiting M., Madan A., Young A.C., Shcherchenko Y., Bouffard G.G.,
RA Blakesley R.M., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.W., Buttefield Y.S.,
RA Krzywiński M.I., Skalska U., Smalins D.E., Scherch A., Schein J.B.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."

```

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN (2)  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J TISSUE=Brain;  
 RA Strauberg R.;  
 RA Submitted (SEP-2003) to the EMBL/GenBank/DBJ database.  
 DR EMBL; BC058729; AAHS8729.1; -  
 FT NON TER  
 FT SEQUENCE 2003 AA; 226929 MM; A130BC7AE053D27 CRC64;  
 Query Match 3.2%; Score 291.5; DB 2; Length 2003;  
 Best Local Similarity 19.2%; Pred. No. 1e-09;  
 Matches 354; Conservative 250; Mismatches 657; Indels 583; Gaps 83;  
 7 TSVVSPGSSRRLWTL---LGRSALRELSQIEALNKHRRLLGSLVYK--PPSPSS 61  
 2 VNSAASLWGPYKOIWTGVSALMRROPEAVHLLDMILKCH--KDPFISLFPKPPKVVQ 58  
 62 AEKV-KANDVASPLKEGLR-----ISKFLGDEBSQVQLQC--YIQEDYR 106  
 59 HEKIQASSEGVAIQGGQTRLLPEGLKEAPFISLPDIGELSANVELLAGHQPHRP 118  
 107 G-TRDSVKTVDODEROSQALLKXADYYEERTCIACVHLITTFQDER-HPTRYEYAD 164  
 119 GLTRGLVAVTL-----YMDGKACIN--SLRTLIQSRGKTWLE--- 156  
 165 CVDKLEKELVSKYRQOFBEELYKTEAPTWETHGNLMEROVSRWFOCCEQSMLEITFL 224  
 157 ---LSPELVSS-----MTR-----FTDELMEOGLTYKVLTL 184  
 225 YVAEYEMAPSDLLV---LTKMFEQGFSGRQTNRLVD-----ETM----- 262  
 185 L-----SQIDVNNPEFKLQREGLGSEKHKRKEVSDLIKRCROSLAESLAPMAQSP 236  
 263 --DPEVDKIGYSALIVEGMDIESLHKCALDREHLPADGLICOD--MDCIMTFE 317  
 237 PKDDTLILIGHLERYVEANGSIDAVNLCL-----NALYCEPFTSFIDSTERR 286  
 318 GDIPHPAPV-----LAWALLRHILN--PE-----E 341  
 287 DDMIHLEPILTERQYVSTHSLRQDSQPKLPGLQATVRLANALRGISQLPDVTALAE 346  
 342 TSSVVRKIGGTAIQIANVFOYLTLLQSLASGNDCTTACMCVYGLSFLVT--SLBIA 399  
 347 FTEADBAIAEAIADNVFLFS---EAVVLAENFQEBEYIRIRHISLIDFLAFAPMKYK 403  
 400 TLGNQD-----IIDRACEVLADPSLP-----ELFWGTE 428  
 404 QLKRADEDARMIHMSIQMGNEPISLRRLDEHMLLIGELYKNPFIHELALBYWCPSBE 463  
 429 FT-----SGLG-----IIDSVCGMFPHLS-----PLIQLLRALVSGSTAK 466  
 464 PLQGTWGSYIGVAAHQRPQROGVLSKVRQMGDLIPPTIYIFLKLQSLANPQCAN 523  
 467 KYVSELDKMSFYNELYKHKPHDVISHEDGTLWRQTPKLLVPLGGQTNLRIPQGTVGVM 526  
 524 YCFSL-----KVNSSHVENIQAGGSP- 547  
 527 LDDRVLVWMEYSVSWTLFTCEIEMLHVSTADVIGHQCVKRIIDLVHNVISTDLSI 586  
 548 -----VSWEFFPSLLLYH--EHLRKDLPSSASVOY--EHLPSRGTQK--BODGI 593  
 587 ADCLLPITSRIYMLQ--RLTIVISP--PVDVIASVNC-----LTVLAA--RN 629  
 594 A--PLQGTSTIITWENARLALCHRPQWTPVVIIIGLQCSIPYLKAEMLKTLTAAFGS 651  
 630 P--AKVWTDLRHTGFLPFAVAVSSLSQMSIABGKNNAGYNNLLMNSQPOGEXVTIA 686  
 652 PEIAASLWQSLLEYTQILQTVRVRSQRAIGIEVE-----LNEIESRCEEPFLTRA 701  
 687 FLRLITLVKQLGSTQSGL-----VPCVMVLKEMLPSTYKTKMYNSHGREQJGCLIL 741  
 702 FCQLISTIVSSFPFNLAGLAPPGPDPYLOFLARDSVTLRFRTAAYRRAAEKMEVAVYL 761

QY 742 ELIHAIINLCHETDHSHTPSL-----QFLCICS---LAYTAG-----Q 779  
 DB 762 EVFYTLR-----DYEQLDEDFVDQFVELQGBEIIAYNPPGFSIMYHLLNSPMLB 812  
 QY 780 TVININGIGVDPIIDWMAAPSPSDAEQGGQGLIKTYKVL-AFAYTNVVRILKPPSNV 838  
 DB 813 LALSLIBSGVQLD--TYAPFPQKGLBEKAVGHCLALNLTLOKENLPMDLRBSQALIV 871  
 QY 839 SPLBQAL---SQHGANNLIVAKTYIYHKIDPALPRIALQLRALATVAPMS----- 889  
 DB 872 SPLBQLQGINPRTKKADNVVNI-ARYLYHGNN--PEIAFESAKITLCISGNSIQVMA 928  
 QY 890 -----VYACIGNDAA--IRDAFLTLQSKIDMR--IKMIIIEPLTV 928  
 DB 929 VQDFTHDSVSGKLMAGVECELDYEDTEFPAVBEGBSELEKKAARHSTRHILNLTIT 988  
 QY 929 AVETOPGLIELP-LNLBYKDSGDSKESFLGMW-----SCLHAVELIDSOQDR----- 977  
 DB 989 SLERNPRLALYLLGFELKPIISTNNLDPPVLAGCPRTCLNAILNLEKGTGRGDPVAV 1048  
 QY 978 --YWCPLIHPAALAFHAL--WQDRDSANLVLTREKFMENLTSPFGLTSP--PSE 1030  
 DB 1049 RBY--POLADLCYQVITQACSPDSGPTRYRTYSODF-----LFSQLOHLFPBNK 1098  
 QY 1031 TSEPSILETCALIMKIICLETIYVYVKGSLDSIMOTLKKPSI-----EKRPAYMSGYK 1084  
 DB 1099 EHEISMLQMSWMLKTSIEIRVT--SLNQRSHQRLHLLDMDVFKPSIDGEGW- 1154  
 QY 1085 SLAVHVAETEGSSCTSLLEYQMLVSAMWMLIITTHADIMHLDVSV-----VRRQLFLD 1139  
 DB 1155 -----EDENRSYSGFLHFDATATVRKILISI-----LDSIDFSGEIRPQLD 1197  
 QY 1140 VLDGKALLVPAVYNCIR--LGSMTLLIILRQKRELGSVDLGPITLIEVL 1196  
 DB 1198 FFDRAQIQOVI--ANCHENKIQOQVYCNVXL-----HRLVAVRNALQWVA 1242  
 QY 1197 QADQOLEKTKAKVFSAPIT-----VLOMKENKVSQIPQYSQVLYNVCSTLQEBVIALPQ 1252  
 DB 1243 AIGQRFILMEISTILLQYVGRANKLQCLAKKRAHBSKQVLVILITACQBELIQADR 1302  
 QY 1253 TRHSIATGASATEDKDSMETDSCSRSHRDQDGVCGVGLMLAKELCEVDEBDGSMLOVTR 1312  
 DB 1303 ---QIIRIDLQVHDVKLDD-----BAQSL----- 1326  
 QY 1313 RLPIPLITTLTLEYSLSKQKYLHTEBATHLLTLARQQAATVANGITQISCLPLIS 1372  
 DB 1327 -MPVAVAGAVFTL-----TAHLISQA-----VRTBORQPLVSGGEAQ-----Y 1362  
 QY 1373 VYOLSTNGTAQTPASR-----KSLD--APSMGVYRLMSIMBQLK 1413  
 DB 1363 AFMDSLSITSSPAESRPVGRASIGDSSIMIIILKKLIDFLIKTGGPQPRVATHLYGSILY 1422  
 QY 1414 TLRNPLPEALDYGVOERTLQCLANAVRYQS-----LACLEBAD 1454  
 DB 1423 YLQIAQRPDEBDTEAAKKTWMBERTTAPBEDVFSKLORENNAAIISYGALAMEVCRCADCD 1482  
 QY 1455 -HTVGFILOLS-----NFMKEMHFHLPOLAMRDIVNIGYVLOACGSLHSKRMLOHY 1505  
 DB 1483 GHEIGRMALALDLRIYSVDQKQOMLYLS-----NSGQYLVYDSLVDDDDRTLOSL 1534  
 QY 1506 LQNKRGDGLPSAVQAVQRPSPSAASAAPSSKQPADTEASEQ 1549  
 DB 1555 L-----TPQPLRLKALTYTBSKMAFLTRVAKEQ 1563

RESULT 18  
 AAHS8729  
 ID AAHS8729 PRELIMINARY, PRT, 2003 AA.  
 AC AAHS8729;  
 DT 02-MAR-2004 (Tribble, 27, Created)  
 DT 02-MAR-2004 (Tribble, 27, Last sequence update)  
 DT 02-MAR-2004 (Tribble, 27, Last annotation update)



QY 1313 RLPILPILTLTLEVSILRMKNLFTETXHTLTLTARTOQATRAVAGITGOSICLPELIS 1372  
DB 1327 -MVAAGAVFTL-----TALHQA-----VTEQROPLVSGPQAQ-----Y 1362  
QY 1373 VYQSTNGTQATPSASR-----KSLD--APSMGYRLSMSGLEQLK 1413  
DB 1363 APFLDSLTSSPAABSRPVGFASIGDSSLHIIILKGLDPLFKTGGRQVRTHLYGSLY 1422  
QY 1414 TLRNPLPEALDPVGVHQBETLOCLNAVRTQS-----LACLEAD 1454  
DB 1423 YQIAORPDEPDTLEAKKTWERTLTAPEDVFSKQRENNALIESYGALMEVVCSDACD 1482  
QY 1455 -HTVGFILDS-----NFMKWHPLPQLWRDQVNLGYLCOACTSLHSRKLQHY 1505  
DB 1483 GHEIGKMLALALIDRIVSVKQHWMLYLS-----NSGYLKLVLVSDLDRLTQSL 1534  
QY 1506 LQNKNGDGLPSAVAGVORPPSPASAPSSSKQPADTASQO 1549  
DB 1535 L-----TPQPLKALYTESKAPILTRAKEQ 1563

RESULT 19  
ID 06ZOG1 PRELIMINARY; PRT; 2067 AA.  
AC 06ZOG1;  
DT 05-JUL-2004 (TEMBLrel. 27, Created)  
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)  
DE 05-JUL-2004 (TEMBLrel. 27, Last annotation update)  
DN MKIA0225 protein (Fragment).  
GN Name:MKIA0225;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OC NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSE=Embryonic tail;  
RX PubMed=14621295;  
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,  
Saga Y., Nagase T., Ohara O., Koga H.;  
RT Prediction of the coding sequences of mouse homologues of KIAA gene:  
RT I11, the complete nucleotide sequences of 500 mouse KIAA-homologous  
RT cDNAs identified by screening of terminal sequences of cDNA clones  
RT randomly sampled from size-fractionated libraries.";  
RL DNA Res. 10:167-180(2003).  
DR EMBL, AK129093; BAC97903.1; -.  
FT NON TER 1  
SQ SEQUENCE 2067 AA; 233812 MW; 84C09D69FD7B0B CRC64;

Query Match 3.2%; Score 289.5; DB 2; Length 2067;  
Best Local Similarity 19.2%; Pred. No. 1.5e-09;  
Matches 356; Conservative 244; Mismatches 645; Indels 609; Gaps 83;

QY 26 GRSALRELSQIEAEI-----NKWRRLLEG-----LSY 53  
DB 53 GRSALSVLLDQAGVGAASIMGPYDQIMQTVGSALMRQPEAVHLLDMILKKHKRDPISL 112  
QY 54 YK--PPSSSAKV-KANKDVASPLKEGLR-----ISKFGLDSEGSVQLIQ 98  
DB 113 FKNPKKNVQOHEKIKKASSEGAIVQOQGTIRLLPEQLIKAFIISDLFDIGELSAVELLL 172  
QY .99 C--YQGEDYRG-TRDSVKTVLQDEROSQALIKIADYYEERTCLRCVHLLTYFOER 155  
DB 173 AGEHQQHPFGILTRGLAVVL-----TWGKRCTAN--SLRTLIQRR 213  
QY 156 -HPYRVEYADCVLKELELVSKYRQOFELXYKTEAPWETHGNLMTTEROVSHWFQCLRE 214  
DB 214 GKWTLE-----LSPELVS-----MTTR-----FTDELME 238  
QY 215 QSMLEITFLYYAYFEMAPSDLLV--LTKFKKQGGFSGRQTNRLVD-----ET 261  
DB 239 QGLTYKVTLL-----SQIDVNNBEKIQRBGRGLSEKIKREVSDDLKRCROSLAS 290

QY 262 M-----DPVDRIGYFSALLVEGMDISLAKCALDDBRRLHQAODGLICOD- 309  
DB 291 LPAWACQSPLPKDDTLILGHLERYTVENGSLDANVLCIL-----NALLYCPDT 340  
QY 310 --MDCLMLTRGDIPIHAPVL-----LAAALLRHTAN 338  
DB 341 SFIDOSTERDDMIHPLTLTRQYVSTHSFLQDSQFWKLGLOATVRLAVALARGIS 400  
QY 339 --PE-----ETSSVVRKIGGTALIQLVNVPYLRLLQSLASGANDCTTSACVCVGLISF 391  
DB 401 QPDDYALABFEADBAIALADNVFLFS--BAVLAANFQGEERYIRHISLIND 457  
QY 392 VLT--SLEHTYGNOD-----IIDRACEVLDPILP----- 421  
DB 456 FLAFNPMKVKQKQKNADBDARMIHNSIQWGNBPISLRDLRHLMLIGELYKNPFILE 517  
QY 422 ---ELFWGTRP-----SGLG-----IILDSVCGMF-----PHLSPLIQLR 456  
DB 518 IALBYWCSEBPLQPTTWGSLGVANORPPORQVLSKFRQMGDLPEPTIIPYIKMLQ 577  
QY 457 ALVSKSTAKKVSFLDKMSFYNELYKHPDVLSHEDGLMRQTPKLYPLAGQTNIR 516  
DB 578 GLANGPQCAHYCFSL-----KVNSSHYE 602  
QY 517 IPQGVGVQMLDDRAVLYRWBYSSWTLPFCIEIMLHVSTADVIHQORVKPTIDLV 576  
DB 603 NIQAGGSP-----VSWHFHSLLLYH--HILKRDLPSSADSVQY--RHLPSRGIT 649  
QY 577 HKVISTDISIADCLPITSRIYMLQ--RLATVISP--PVDVIVSCNVC----- 621  
DB 650 QK--QDGLIA--FLQISTITISRNARLALCHRPQTPPVVILGLQCSIPPLKAE 705  
QY 622 LVTAA--RNP--AKVWTDLRHGFPLPVAVPVSSISQMTSABGMNAGYGNLLMNSQ 676  
DB 706 LKTLAFAKSPBIAASLWQSLLEYQILQTVAVPSQRAIGIEV-----LNEIES 755  
QY 677 POGEGVIAFLRLITTLVKQLOGSTQGL-----VCWAFVLEKMLPSYHKRYNSHG 731  
DB 756 RCEBPLTRACQLSLTVSPPSNTGAGLRPGPDFYLOPLRDSVPLRFTTRAYRBA 815  
QY 732 VREQIGCLILHAIHILNCHEIDLSHTPSL-----QPLCIS--LAYTEAG----- 778  
DB 816 EKMEYAEVVLVFFKLR-----DYEQLEBPVDQFVBLQSEBIIAYKPPPSLMY 866  
QY 779 -----QVINIMIGVDITDMVMAAOPRSDAGEGGQGLIKTVKL-AFSVTNNV 828  
DB 867 HLINESPMLBIALSLBERGVQLD--TYAPFGKHLKRAVOHCLALINLTLQKENVL 925  
QY 829 IRLKPPSVVSLQAL-----SQHAGNMLAVLAKYIYHGHDPALPRLAQLKRLATV 885  
DB 926 LBSQGLALIVSLBQLOLQGINPRTKADNVNI--ARYLYHNNN--PELAESAACILCCI 982  
QY 886 APMS-----VYACLGNDAA--TRDAFLTLQSKIEDMR--I 918  
DB 983 SCNNIQVQWVGDPETHQSVQKLMAGVBECLDVBDEBFPRVBEGBSELEKKAIRRET 1042  
QY 919 KVMLEFLTAAVETQGLIEIF-LNLEVQSGDSKESFSLGM-----SCHAVLELDSQ 973  
DB 1043 RIHITIMLITSLERNPMLALYLLGFELKPIISTNNLODPVLACPCFRTCAIILNILEKG 1102  
QY 974 QODR-----WCPPLRHAATAFLHAL--WODRDSANLVLRKPKFENLTSPLFG 1023  
DB 1103 TEGRDGPAVAREY--POLADLCYQVYIQLACGSDTSGPTRYKRYLTSODF-----LFS 1152  
QY 1024 TLSP---PSRTSEPSILETCALIMKIICLEIYVVVKSGLDQSLKDTLKKFSI-----EK 1074  
DB 1153 QIQHLPPSNKKEHISLMSQSWLKTASIEHRYV--SIARQSRHTQRLHLLDMDMPVK 1209  
QY 1075 RPAVWSGYKASLAVHVAETGSGCTSLLEYQMTLVANRMLIITATTHADIMHLDTSV-- 1131  
DB 1210 PYSDBGEGM-----EDKRRSVSGFLAHPDTATKVRKILISI-----LDSIDFS 1251

QY 1132 --VEROLFVLDGDKALLLVPAVNCUR---LGSMDCTLLILDRQKRELGSDVDEILG 1186  
 DB 1252 QEIEPFLDLPFDRAQIQV---ANCENKLOGTVCNKL-----HRLV 1296  
 QY 1187 PLTEILEGVLADQOLMEKTRAKVFAIT---VLQKMEKVSIPQYSQVILNVCETL 1242  
 DB 1297 AEVVALQGMALIGORFLMEEISTILQYVGNKLLQCLHAKRHLESRMQVBEIILTAC 1356  
 QY 1243 QEEVIALFDOTRHSLALGSATEDKDSMETDCCSRNRHDDQDGVVGLHAKELCEVDE 1302  
 DB 1357 POELIOAEDR---OLIRDLQDVADKVLDD-----EAAQEL----- 1390  
 QY 1303 DGDGMLQVTRRLPLPLTLTLEVSIRMKONLHFEATLHLLTLARQOGATAVAGAI 1362  
 DB 1391 -----MPVAGAVFTL-----TALISQ-----VTEKORPLVSGFGE 1423  
 QY 1363 TOSICLPULSVYOLSTNGTAOTPSAR-----KSLD--APSMGVYRL 1403  
 DB 1424 AQY-----AFMLDSSLTSPASBRPVGFASIGDSSLHILKLLDPLIKTGGGQQRV 1476  
 QY 1404 SMSLMELKLTNRNPLPEALDFVGVHOFRTLOCLNAVTVOS----- 1446  
 DB 1477 RTHYGLSLVYVQLAQRPDEEDTLEAKKTWERLTAEDVFSKLORENMAIIBSYGAL 1536  
 QY 1447 --LACLEBAD-HTYGFILOLS-----NPKKHFFLPOLMRDIOVNLGLOACTSL 1495  
 DB 1537 MEVVCRDACDGEHIGRMALALDLRIVSDVQOHQWLYLS-----NSGLKTVLDSL 1588  
 QY 1496 LHSKMLQHYLQNKNGDGLPSAVAQRVORPSPASASAPSSSKOPADTEASEEQ 1549  
 DB 1589 VDDRTIQLSL-----TPQPLKALVYTESGMARLTVAEEQ 1627

## RESULT 20

BAC97903

PRELIMINARY, PRT, 2067 AA.

BAC97903

BAC97903

02-MAR-2004 (TREMBlrel. 27, Created)

02-MAR-2004 (TREMBlrel. 27, Last sequence update)

02-MAR-2004 (TREMBlrel. 27, Last annotation update)

MKIAA0225 protein (Fragment).

MKIAA0225

Mus musculus (Mouse).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

[1]

SEQUENCE FROM N.A.

RC TISSUE=Embryonic tail;

RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hirooka S.,

Saga Y., Nagase T., Ohara O., Koga H.,

"Prediction of the Coding Sequences of Mouse Homologues of KIA Gene:

RT i. The Complete Nucleotide Sequences of 500 Mouse KIAA-homologues

cDNAs Identified by Screening of Terminal Sequences of cDNA Clones

Randomly Sampled from Size-fractionated Libraries."

RL DNA Res. 10:167-180(2003).

DR EMBL; AK129093; BAC97903.1; -

FT NON TER 1 1

SQ SEQUENCE 2067 AA, 233812 MW, 84C09D69FD77B0E CRC64;

Query Match 3.2%; Score 289.5; DB 2; Length 2067;  
 Best Local Similarity 19.2%; Pred. No. 1.5e-09;  
 Matches 356; Conservative 244; Mismatches 645; Indels 609; Gaps 83;

QY 26 GRSALRELSQIABE-----NKMRRLBS-----LSY 53  
 DB 53 GRSALRELSQIABE-----NKMRRLBS-----LSY 53  
 QY 54 YK--PPSPSAKLV--KANKDVASPLKELGLR-----ISKFLGDEEGSVQLIQ 98  
 DB 113 FKNPFPKVVQGHKIKQKASBGVALIGQGGTLLPQLIKBAKATISDLPDIGLSAVELL 172  
 QY 99 C--VLQEDYRG--TRDSVKTVLQDERQSAALIKIADVYTERTCILRCVLIHLITTFQDBR 155

DB 173 AGEHQDHPFGLTGLVAVLL-----YMDKRCIAN---SLRTLIQERR 213  
 QY 156 -HPYREYADCVQKLEKELYSKTRQOEELKYTAPTWETHGNLMTREOVSRWVQCLRE 214  
 DB 214 GKTWTLR-----LSBELVS-----MTTR-----PTDELMR 238  
 QY 215 QSMLEIIFLYVAFFENAPSDLV---LTKMFKEGSGSRQTNHLYV-----ET 261  
 DB 239 QGLTYKVLTLL-----SQTVDNNEPEKQREKRGKSKKRVSDLIKECROSLAES 290  
 QY 262 M-----DEFVDRIGYFSAIIVEGMDISLKCALDRREHQAQDGLICOD- 309  
 DB 291 LFMAAGSPLEKDTLLIGHLEKRYTAEANGSLAVNLCIL-----MALLYCPDT 340  
 QY 310 -MDCLMTFQDIDPHHAPVL-----LMAILRLTLN 338  
 DB 341 SFIDQSTEERDDMTNHLPLTERQVSTHISRLQDSQPKWLGLOATVRLMALALRGIS 400  
 QY 339 --PE-----ETSSVVRKIGTAIOLANFOVYTRILQSLASGNDCTTACKCVYGLISF 391  
 DB 401 QLPVYTALEFTEADEALIALADNVFLFS---EAVVLAENFYQEEFYTRIRHSLITD 457  
 QY 392 VLT--SLELTAGNOD---IIDTACEVLADPSLP----- 421  
 DB 458 FLAMPKVKVQLKRADEADAMIMSTOMGNRPISLRDLEHMLLIGELYKKRPHLE 517  
 QY 422 --ELFWGTEPT-----SGLG-----IILDSVCGMP-----PHLSPLQLLR 456  
 DB 518 LALEWYCPSEPLQPTWNGSYLGAHQRPQQRVLLSKFVQWMDLLPPTIYIPLYKQLQ 577  
 QY 457 ALVSGKSTAKKRVYFLDKMSFYNELYKHKRHVDVSHEDGLMRQTPRLYPLQGGTMR 516  
 DB 578 GLANGPOCAHYCFSL-----KVGSSHVE 602  
 QY 517 IPQGVQVMDDAVYLVREWEYSYSWTLFTCEIEMLHVSTADVIOHCORVXPIDLV 576  
 DB 603 NIQAGGSP-----VSWHFPHSLILYH---EHLKDLPSADSVQY--RHLPBRI 649  
 QY 577 HKVISTDLISADCLPITSRIYMLQ--RLTVIAP--PVDVIASCVCN----- 621  
 DB 650 QK--EODGLIA--FLQITSTIITWSENARLALCHHPQWTPVVVILGLQCSIPLYKABL 705  
 QY 622 LTVLAA--RNP---AKWTDLRHGTPRPVNAHPVSSLSQMTSAGMANAGGNLMSQ 676  
 DB 706 LKTLAAREKSPBIASIMQSLIYQIQTAVVPSQROAIGIEV-----LNEIBS 755  
 QY 677 PGEYGVYIAFLRLITLVKGLQSGTOSQGL-----VPCVWFVLEKMLPSYHKVRVNSHG 731  
 DB 756 RCSEYPLTRAFQOLISTLVESPPSNIGAGLRPGFPDYLQFLDSDVFLRRTAAYRGA 815  
 QY 732 VREOIGCLIELHAIINLCHRTDLHSHRPSL-----QPLCIS--LAYTAG----- 778  
 DB 816 EKMEVAEVLVEVFKLKR-----DYEQQLDVPDQVFELOEELIAVKPQGFSLMY 866  
 QY 779 -----QTVINMIGIVDTIMMAAQRSGASGQGGOLIKTVK--ASVTVNV 828  
 DB 867 HLINESPMLBIALDLLEGVQQLD--TYAFPQKHLKEAIVQHCALINTLTQKXNLPFDL 925  
 QY 829 IRLKPPSVNVSLQAL---SQHGAGNNLAVLAKYIYHGDALPRLATQLKRLATV 885  
 DB 926 LRESQALIVSPLQOLQGINPRTKADNVNT--ARTVYHGNNA--PELASAKILCCT 982  
 QY 886 APMS-----VYACIANDNAA--TRDAFLTRLQKIDMR--I 918  
 DB 983 SCNNINIQVQWGVDFTHQSVSQKLAGVLELDVDETEFVRVEGSELEKKGAIRHET 1042  
 QY 919 KVMILFETVAETOPGLILP--LNLVYQSGDSKESPLGMM-----SCHAVLELIDQ 973  
 DB 1043 RIHILNLTLSLENNPRLALYLLGFLKPKISTTNLQDPVLCOPRYCAHAIINLIEKG 1102  
 QY 974 QODR-----YMCPLRHHAAALAFHLA--WQDRDSAVLVLRKFKWENLTSPLFG 1023

```

Db 1103 TEGRDGPVAVREY---POLADLCYQVITQACASDGTSGPTWRYLRTSGDF-----LFS 1152
Qy 1024 TUSP-----PSRSPSILETCALIMKICELIYVVKSLDQSLKDTAKEST-----EK 1074
Db 1153 QLOHLPSPSKHEHISMLQSMWMTKASISLRYT---SLMRORSHQORLHLHLLDDMPVK 1209
Qy 1075 RFAVWSGVYKSLAVHVAETEGSSCTSLLEYQMLVSAVRMLLIATTHADIMHLLTDSV--- 1131
Db 1210 PYSDGGGGM-----EDENRSVSGFLHFDYATKVRKRLST-----LDSIDBS 1251
Qy 1132 --VRQFLPDLVDLTGALLVLPASVNCLE---LQSKCTLLILLQWKEELGSDVDEIG 1186
Db 1252 QEIPEPLQDFFPRAQIEQVY---ANCEHKLQGOQVYCNVCL-----HRVLY 1296
Qy 1187 PLTEIEGVQADQOLMEKKAIVSFAIT---VLOKEMKSDIPQYSQVLANCETL 1242
Db 1297 AERNALQGMALIQORPLMEISTIIQYVGNKLLQCHAKHAALESWQVLEITLITAC 1356
Qy 1243 QBEVIALPQTRHSLALGATDEKDMETDDCSRSRRHQRDGVGLHLAKELCEVDS 1302
Db 1357 PQLIQAEHR---QLIIRDLQDVHDKVDD-----EAGEL----- 1390
Qy 1303 DQDSMLQVTRRLPILFTLLITLESVLRMKNLHFEATLHLLTLARTQOGATVAVAGAI 1362
Db 1391 -----MFWVAGAVFTL-----TAHLSQA-----VTRQORQPLVSGPGE 1423
Qy 1363 TQSLCLPLSVYQSLNGTGAQTPSASR-----KSLD---ASWMPQVYL 1403
Db 1424 AQY-----AFMLDSSLTSSPAESRPVGFASIGDSSLHLLKCLDLPLKGGGFORV 1476
Qy 1404 SMSLMEQLLKTLYNLPALDFVGVHQBRTLOCLNAVTVOS----- 1446
Db 1477 RTHVGLSLIYVLIQIAPRDEPDTLEAKKTMWERLAPADVPSKQOREMMAIIESYAL 1536
Qy 1447 --LACLEAD-HVTEGTLQSL-----NPKEMHFLPLQMRDIQVNLGTLCOACISL 1495
Db 1537 MEVVRDADCDGHEIGMLLALDLRVSVDKQHWMLYLS-----NSGYLKVLYVSL 1588
Qy 1496 LHSRKLQHYLQNKQDGLPSAVAQVORPPSAASAAPSSKQPADTASQ 1549
Db 1589 VDDRTQLSL-----TPQPLKALITYESKAPLTRVAKEQ 1627

RESULT 21
O6P9L5 PRELIMINARY: PRT: 1851 AA.
ID O6P9L5 AC O6P9L5
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DB Nup205 proteol (Fragment).
GN Name=Nup205;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=1090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Altschul S.F., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carrinck P., Prange C.,
RA Boeak S.A., McEwan N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villation D.K., Muzny K.C., Sodergren B.O., Lu X., Gibbs R.A.,
RA Bailey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

```

```

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalios D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maira M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strauberg R.L.
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strauberg R.L.
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC060711; AAH60711.1; -.
DR EMBL, BC060139; AAH60139.1; -.
FT NON_TER 1
SQ SEQUENCE 1851 AA, 209925 MW, DD11A9847159291 CRC64;

Query Match 3.2%, Score 286, DB 2, Length 1851,
Best Local Similarity 19.2%, Pred. No. 2.1e-09;
Matches 334, Conservative 236, Mismatches 621, Indels 550, Gaps 76;

Qy 15 RSSRLMTILLGRSLRLSOTIEALNKMRLL--BELSTYKPPSPSAEKYKAKOVA 72
Db 15 RFTBELMQGGLTYVTLTSLQI--DVNNEPRLQRRERIG-----SEKIRKEVS 61
Qy 73 SPLKELGLRISKPL-----GLDEEOSVOLLOCLQE--DNGTRDSVYTVLDERQS 122
Db 62 DLNCRQSLAESLPAACOSPPLPRDITL-LIGHLEVTYVANGSLDAV----- 110
Qy 123 QALIKADVYBERTCILRCVHL-L-TYFODERHPYVEYADCVDKL-EKELYSKYRQ 180
Db 111 -----NLGIMALVLCPTSFIDQTEERDDMIHHLPLTERQYVSTHSR 156
Qy 181 FEELYKTAFTWETHGNLMTERQVSRWFVQCLREQSMLEI----- 221
Db 157 LQ-----DSQPKLPGLQATVRLA--MAL--ALRGISQLPVTALAEFTADEAIAELA 208
Qy 222 --IFYAYVFEMADSDLLVTKMFKEOGSGRQNRHLVDETMDFVDRIGYPSALIVE 279
Db 209 DNVEFLF-----SEAVLAKENFYQSEFTYRRLHSITTD-----FLAFM-- 246
Qy 280 GMDIESLHKCALDRLRHQPAQDQ-----LICODMCLMTFGDI-----PHNAPVLLA- 329
Db 247 PMKVKQKNRADBEDARMIHWSIQGNRPPISLRDLRHMLLIGELYKQNPFLIELALEY 306
Qy 330 WALRHNLNPEBSSVYRKIGGTALQIANVFOYLTRLLQSLASGANDCTSTACMCVTGL 389
Db 307 WC-----PSEPLQTPVWGSYLGVA----- 326
Qy 390 SFVLTSLEHLTGQOQDIITPACVLAADPSLPELFWGTEPTSGIITLDSVGMFPHLS 449
Db 327 -----HQRPRQRYVLSKPYQWGDLP-----PTIYT 354
Qy 450 PLOQLRALVSGSTAKVYSLDKMSFYNELYKHKPHDYISHEDGLMRRQTEKLLYPL 509
Db 355 PYLMLQGLANGPQCAHYCSLL-----KV 379
Qy 510 GGGNNMLAPGGTVQVWLDDBRATLYMRVYSQSWTLFTGIBMLHVVSTADYVQHCORV 569
Db 380 NGSSHVENIQAAGSP-----VSWNEHPSHLLTH---EHLRKDLPSADSVQY--RH 426
Qy 570 KPIIDVHKYISTDLADCLLPTTSRYMLQ--RLTVIISP---PVDIASVNC--- 621
Db 427 LPSNGIQK--EDDGLIA--FLQITSTIITWSRNALALCENHPWTVVVILGLQCSIP 482
Qy 622 -----LTVIAA--RNP--AKVWTDLRRTGFLPFVAVHVSLSQWISAEGNAGGYGN 669
Db 483 PVLKAEILKTLAAGKSPBIAAGIMQSLBETTLQVTRVPSQQAIGIVBE----- 533

```



```

QY 670 LMANSEPOGEGYVTAFLALITLVKQGLSTOSQGL-----VPCWFLVAKENLPSYHK 724
DB 534 -LNEIEBRCREYPLTRAFQCLITLVBSFPSNAGAGRPBGDPYLOFLDSEVLEBRT 592
QY 725 WRVNSHVRBOIGGLIELIHAIALNLCHETDLSHSHPSL-----QFACIS--LAYTE 776
DB 593 RAYRAAEKKEVNAEVALEVEFYKLR-----DYEQLQEDPDQFQELQGBEITAYKP 643
QY 777 AG-----QVINIMIGVDTIDMNAAPRSDGAGGQGLITVKL-A 821
DB 644 PGSLWVHLNBSMELALSLBEGVQLD-TYAPFGKGLKRAVCHALALNTLQK 702
QY 822 FSVTNNVIRLKRPNSVNSPLEQAL-----SQHAGNNLAVLAKTYHNDALRLAIQL 878
DB 703 ENLFMDLRBSQALIVSPLQGLQINPRTKADNVNI-ARLYHGNNA--PELAFES 759
QY 879 LKRLATVAPMS-----VYACLGNDAAA--IRDAFLTRLOSKI 913
DB 760 AKILCISCSNIVKAVGDFTHQOSVQKLMAGFVGLDYEDTEBEFVREGESELEKCL 819
QY 914 EDNR--IKVMILEFLTVAVETOPGLIELF-LNLEVKGSDGSKFSLGMW-----SCLHAV 966
DB 820 AAIHREHRIHNLITSLKENPMLALYLLGFELKKRISTTNLODPVLCSPRTCLAI 879
QY 967 LELIDSOQOQR-----YCPPLHRAALAFHLAL--WDRRDSANLVLTREKPEWEN 1016
DB 880 LNLIEKTEBGRDGVAVAREY--POLADLCYOVLYOLCACSDDTSQPTRYLRTSDF--- 933
QY 1017 LITSLPGLTSP--PSETESESLIETCALINKITCLEIYVYVKSGLDLSQTLKKFSI- 1072
DB 934 ---LFSQLOHLPPSNKEHEISMLSQSWLMTKTSIEIRVT--SLNQRSHTORLHL 986
QY 1073 ---EKRFAYSGVYKSLAVHVAETEGSSCTSLLEYOMLVANRMLIATTHADIMHL 1127
DB 987 LDDMPKYPSDGBGM-----EDENRSVSGFHPRTAKVNRKIKLSI----- 1028
QY 1128 TDSV-----VRQLFLVDLDTKALLLVPAVSNCR--LSQKCTLLILLRQKBEIG 1179
DB 1029 LDISDFQEIPEPLQDFPDBAQIEQVI--ANCEHKNLQCYVCNVKL----- 1075
QY 1180 SVDEILPRLTILGCVLQADQGLMEKTKAYPSAFT--VLQKMKVSDIPQYSOLV 1235
DB 1076 -HRLVAEVAVALQGMALIGRPRLMEIISTILQYVGRNKLQCLAKRHALLSWROLV 1133
QY 1236 LNVCELTQOEVIYALPDOTRHSIALGASATEDKDSMETDSCSRSHRDQDGVCLGLHLAK 1295
DB 1134 EITITACQELIOADR--QLITRDLQDVHDKLDD-----EAYQ 1172
QY 1296 ELCEVDEBDSGLQVTRRLPILPTLLTLEVLRLMKONLHTEATVILLTLARTQCAT 1355
DB 1173 EL-----MPVAGAVFTL--TAHLQA-----VTEBQROP 1200
QY 1356 AVAAGGITQSLPLSLVYQSTNCTAOTPSASR-----KSLD--APS 1396
DB 1201 LVSPBGEAQ-----YAFMLDSSLTSSPAPASRVSFGASIGDSSMLIILKKLDPILIKT 1253
QY 1397 WPGVYRLSMSLMEQLLKTLYRNFLEALDFVGHQERTLOCLINAVTVOS----- 1446
DB 1254 GGGFORVATHLYGSLILYLOLAQRDEBPTLEAAKKTMEKRLTAEDVPFSLQREMAAI 1313
QY 1447 ---LACLEBAD-HTVGFILQLS-----NFMKEMHFLPOLMRDIOVNLGYL 1488
DB 1314 ESYGALMEVVCRCADCGHEIGRMALALDRIVGVKOHQWLYLIS-----NSGVL 1365
QY 1489 COACTSLHAKRMLQHYLQNKKGDLPSAVAVQVRPSPASAAASSSSKOPAADTEASEQ 1548
DB 1366 KVLVDSTLVDDBRTQSLT-----TPQPLLKALYTESKMAFLTRVAKQ 1410
QY 1549 Q 1549
DB 1411 Q 1411

```

```

RESULT 22
AAH60139 PRELIMINARY; PRT; 1851 AA.
AC AAH60139;
DT 02-MAR-2004 (Tremblrel, 27, Created)
DT 02-MAR-2004 (Tremblrel, 27, Last sequence update)
DT 02-MAR-2004 (Tremblrel, 27, Last annotation update)
DE NUP205 protein (Fragment).
GN NUP205.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
XP [1]
XP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22386257; PubMed=12477932;
RA Strauberg R.L., Peingold B.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
RA Datchenko I., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Prange C.,
RA Brownstein M.J., Uedini T.B., Tothiyaki S., Carninci P.,
RA Raha S.S., Loguelfano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Borak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green B.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzywnski M.I., Skalske U., Smalins D.B., Scheraga A., Schein J.B.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences".
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strauberg R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC060139; AAH60139.1; -.
FT NON_TER
FT SEQUENCE 1851 AA; 209925 MW; DDF11A9847159291 CRC64;
Query Match 3.2%; Score 286; DB 2; Length 1851;
Best Local Similarity 19.2%; Pred. No. 2,1e-09;
Matches 334; Conservative 236; Mismatches 621; Indels 550; Gaps 76;
QY 15 RSSRBLWTILGRSALRELQIEALNKHWRLL--EGLSYKPPSPSAAKRVANKOVA 72
DB 15 RFTBELMEQGLTYKVLTLISQI--DVNNBEFKLORERGLG-----SEKIRKEVS 61
QY 73 SPLKSLGIRIKEL-----GLDEBSVOLLOCYLQD--DYRGTRDSVYTVQDERQS 122
DB 62 DLIRECROSLAESFPANACGSPLEKDDTL-LHGLELVTVANGSLDAV----- 110
QY 123 QALILKADYAYBRTCLIRCVLHL--TYQDERHPYVAVYADCVDKL-EKELYSKYRQ 180
DB 111 -----NLCILMALVLCPTSPFDQSTEBEDMDIHHILPILTERQYVSTHSR 156
QY 181 FEELYKTEAPTWETHGNLMTREVQSRMFVQCLREOSMLBT----- 221
DB 157 LQ-----DSQPMKLPGLOATVRLA--WAL-ALRGISQPLDVTALAEPTADEAIAELAIA 208
QY 222 -IPLYYVFEAASDULVLTMKRQGFSGRQTNRLHVDSTMPFVDRIGVFSLILVB 279
DB 209 DNVPLFTL-----SEAVVLAENFYQBEFYIRRHSLITD-----FLAFM--- 246
QY 280 GMDTSLHKKALDDRRLEHQAQDG-----LICODMCLMTTFGDI--PHHAPVLLA- 329

```

```

Db      247 PMKVQKLNKRNADDEADAMHMSIQMGNEPPISLRDLNMLLISELYKMPFHELEALEY 306
Qy      330 WALKRHTLNPEETSSVAKIGGTALQNLVFOYLTRLQSLASGNDCTTACMCVYGLL 389
Db      307 WC-----PSEPIQOTPTVMGSYLGA----- 326
Qy      390 SFVLTSLEHLNMGQODIIDPACVADPSLPELPMCTEPTSGIGTILDSVCKMFPLLS 449
Db      327 -----HOPRQROVAVLSKFRQMGDLPL-----PITYI 354
Qy      450 PLLQLRALVSGSKTAKKVSFLDKMSFYNELYKHPHVDYSHEDGYLMRQTPKLLYPL 509
Db      355 PYLMLQGLANGPCAHYCFSL-----KV 379
Qy      510 GCGTNLAPGCTGQVWLDRAVLYVWEYSWTLETCIEMLHNVSTADYIOHCQRY 569
Db      380 NGSSEHNIGQAGSP-----VSEHFFSHLLTH-----EHLRKLPSADVQY--RH 426
Qy      570 KPIIDVHKVISTDLSTADCLPITSRIMLQ--RLTVISP--PVDYIACVNC-- 621
Db      427 LPSRGITQK--EGDGLIA--FLQITSTITWSENALALCEHPQMTPVVITIGLQCSIP 482
Qy      622 -----LTVLAA--RNP--AKVWTDLRHTGFLPFVAFVSSLSQMSAEGMAGYGN 669
Db      483 PVLKAEILKTLAFAKSPETIASLSMQSLLEYTOILQTVRVPDSRQALGIEVE----- 533
Qy      670 LKNSBPQCEYGVTAFLRLITLYKQGLSTQSGL-----VPCWFLKEMLSYHK 724
Db      534 -LNEISRCEBYPLTRAFCOLISTLVSSPFLAGLPPGDFPLQGLRDSVFLRFRT 592
Qy      725 WRNSHGVBEIGICLLLELHAILNCHETDLSHTPSL-----QFLTCS--LAYPE 776
Db      593 RAYRAAEKWEVAEVLLEVFYKLR-----DEPDEPVPVQFVLEQEBEIIAYRP 643
Qy      777 AG-----QVINMGIGVDTIDMMAAPPSDAEGGCGQLIKTVK--A 821
Db      644 PGEISMTLLHNSPMLELALSLEBVGKQID--TYAPFPCKKLEKAVQCHLALMLTLQK 702
Qy      822 FSVTNVAVILKEPSPNVVSPLEQAL--SOHGANNNLIVLAKTYHKGDPALPRLAQ 878
Db      703 ENLFFDLRESQALIVSPLEOLGQINRTKADNVVNI--KRYLPHGNN--PELAFES 759
Qy      879 LKRLATVAMS-----VYACIGDAAA--IRDAFLTLQSKI 913
Db      760 AKITLCISGNSNIQVAVVDFTHDQSVQKLMAGFEGLDVEDTEFVRVEGSLKXKL 819
Qy      914 BMR--IKMILEPLTVAVEQPGILIEP--LMEVYKQSDSGSKESLGMW--SCHAV 966
Db      820 AAIRIETRIHILMLITSLENNPVLALYLLGFELKPISTTNLODPGLGCPRTCLHAI 879
Qy      967 LELIDSOODR-----YMCPRLLHRAAIAFLHAL--WODRDSAMLVLRTPKPMEN 1016
Db      880 LNIIEKTEGRDGPVAVREY--POLADLCVQIVTOLACBSPTSGPTKRYRTISODP-- 933
Qy      1017 LTPLEFGLTSP--PSETSEPSILETCALINKIICLEIYVYVKGSLDLSKDTLKRSI- 1072
Db      934 ----LPSQLQHLPSNKEHEISMLSQMSMLKMTASIEIRVY--SLNRQSHQTRLHL 986
Qy      1073 -----EKRFAYWSGVKSLAVHVAETEGSSCTSLLEYQMLVSNARMMLIITTTADIMHL 1127
Db      987 LDMFVVKYSDGEGM-----EDENRSVSGFHFDTAKVRKIKLSTI----- 1028
Qy      1128 TDSV-----VARQFLVDLGDTKALLVPAVNCIR--LQSMKCTLLILLRQMKELG 1179
Db      1029 LDSIDFQOEIPEPRLQDFPDAQIEOYL--ANCEHKNLQGYTVANVKL----- 1075
Qy      1180 SVDEIIEPILIEGVQADQDLMEKTKAKYFSAPIT--VLOMKENVSDLPQYSQV 1235
Db      1076 --HRVLAEVNVAQGMALIGRPLIMEISTILQYVGRNKQLQCLAKKHALESRWQV 1133
Qy      1236 LNVCEIIEVIALFDQTRHSIALGSATPDKSMETDCCSRRRDORDQVGLGLHLAK 1295
Db      1134 EIIITACPOGELIQMEDR--QLIRDLQDVHNDVLD-----EAAQ 1172

```

```

Qy      1296 ELCEVEDGDSWLVQTRRLPIPLITLTTFVSLRKNQHLHTEATLHLTLARQCGAT 1355
Db      1173 EL-----MPPVAGAVFTL-----TAHLSQA-----VRTEQROP 1200
Qy      1356 AVAGAGITQSIGLPLLSVYQSTNGTAQTPSASR-----KSLD--APS 1396
Db      1201 LVSGFGBAQ-----YAFMLDSBLTSSPAESRVRVGRASIGDSSLIHLIKGLDPLIKT 1253
Qy      1397 WGVYVRLNSLSMLEQLKTLRNFLEBALDFVGHQERTLOCLINAVTVQS----- 1446
Db      1254 GGGFQGVATHLYGSLLYLTLQIAQRDEPDTLEAAKTMERLTVAPEDEVFSKLQRENMAII 1313
Qy      1447 -----LACTERAD--HTVGFILQLS-----NPMKMHPLPOLMDIQVNGYL 1488
Db      1314 BSYGAALMEVCRDCCDCHIEGRMLAALLRLRIVSVQKHQMLYLS-----NSGYL 1365
Qy      1489 COACTSLHSKRMLOHTLQNTKNGDGLPSAVQVRQPPSAASAAPSSKQCPADTEASBQ 1548
Db      1366 KVLVDSLVDDDRITLQSLT-----TPQPLKALTYSSKMAFLTRVAKQ 1410
Qy      1549 Q 1549
Db      1411 Q 1411

RESULT 23
AAH60711
ID AAH60711 PRELIMINARY, PRT, 1851 AA.
AC AAH60711;
DT 02-MAR-2004 (TRENBLrel. 27, Created)
DT 02-MAR-2004 (TRENBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TRENBLrel. 27, Last annotation update)
DE NUP205 protein (Fragment).
GN NUP205.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC NCBI_TaxID=10090;
LN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINER=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshyuk S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bobak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richardson S., Worley K.C., Hale S., Garcia A.W., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman A., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko I., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green B.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyviński M.I., Skalska U., Smalhus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBS databases.
DR EMBL; BC060711; AAH60711.1; -.
FT NON_TER
SQ SEQUENCE 1851 AA; 209925 MW; DDF11A9847159291 CRC64;

Query Match 3.2%; Score 286; DB 2; Length 1851;

```



Beet Local Similarity 19.2%; Pred. No. 2.1e-09;  
Matches 334; Conservative 236; Mismatches 621; Indels 550; Gaps 76;

QY 15 RSSEBELTILGRSLRELISOIEALNKRRL--EGLSTYKPPSPSAEKYANKOVA 72  
DB 15 RFTBELMEQGLTYKVLTLISQI--DVNNEPEKLOREGIG-----SEKRRKVS 61  
QY 73 SPLBELGIRISKFI-----GLDEBSVOLLOCTLOE---DYRGTRDSVKTVALDERQS 122  
DB 62 DLIRECROSLAESLEPAMACOSPLPDDTL-LTGLHERVTEANOSLDV----- 110  
QY 123 QALIKLADYYEERTCILRCVTLHL--TYFODERHPRYEYVADCVKL-EKELSKYRQ 180  
DB 111 -----NLCILMALLYCFDTSFIDOSTEEBDDMIHHLPLTERQYVSTIHSR 156  
QY 181 FEELYKEAPFWETHGNLINTERQVSRWFVOCIREOSMLEI----- 221  
DB 157 LQ-----DSQPMKLPGLQATVLA--MAL-ALRGISQLPDVTALAEFEADEAIALA 208  
QY 222 --ILYVAYFEMAPBDLLVTMKFEGOGFGRQTRHLVDEMDPFVDRIGFSLIIVE 279  
DB 209 DNVEFL-----SEAVVLAENFYOESEFYIRIHSITD-----FLAFM-- 246  
QY 280 GMDIESLHKALDDRELHOFADG-----LICQMDCLMLTFGDI---PHHAFVLA- 329  
DB 247 PMKQKOLKNRDEBARMHMSIQNGNEPPISRDLHEMLLIGLXKPNPHLEALEY 306  
QY 330 WALLRHTLNPETSVMRKIGGTALQNLVFOYLTLLLOSLASGNDCTTSTACMCVGL 389  
DB 307 WC-----PSEFLOTPTVMGSYLGA----- 326  
QY 390 SFVLTSLEHLGNODIIDTRACVTLADPSLPELFWGTEPISGLIIDSVCGMPHLS 449  
DB 327 -----HRRPPQRQVLSKFRWOMDLP-----PTIYI 354  
QY 450 PLCLRLALVSGSTAKKVSFLDQMSFYNELYKHKPHDVISHEDGTLRQTPLYLPL 509  
DB 355 PLYKLOGLANGPOCAHYCFSL----- 379  
QY 510 GQGTNLRIPOGTGVQMLDRAVIVREXSYSSWTLFCEIEMLLAVSTADVIQCORV 569  
DB 380 NGSSVENIIOGAGSP-----VSMHFHSLLYH---EHLKRD.PSADSVQY--RH 426  
QY 570 KPIIDLVHKVISTDLSADCLPTISRIMLIQ--RLTVVSP---PVDVYASVCNC-- 621  
DB 427 LPSRGITOK--EODGLIA--FLOLTSTIITWSENARLALCEHPQWTPVAVVILGLQCSIP 482  
QY 622 -----LTVLAA--RNP---AKWMTDLRHTGFLPFVAPVPSLSQMSABEGMAGYGN 669  
DB 483 PVLKAEELKTLAFAFGSPBIAASIMOSLEIYQILQTVRVSORQALGLEVB----- 533  
QY 670 LMANSEPOGEYGVITIAFLRLITTLVKQGLSSTOSQGL-----VPCVMEVLKEMLPZYHK 724  
DB 534 --LNEIESCEBEPPLTRAFQGLISTLVESFPNLAGARPPGFDDYLGFLRVSFLRFT 592  
QY 725 WRYNHGVREQIGLILBILHAIINLCHETDHSHTPEL-----OFLCTCS--LAYTE 776  
DB 593 RAYRRAAEKWEVAVLEVFYGLR-----DYEPQLEDPVQFVELQSEBETIAYKP 643  
QY 777 AG-----QVININMGIVDTIDVMAAOPRSDAEOGQGLIKTKVK--A 821  
DB 644 PGFSIMYHLNESPMLELALSLLEBVGKQD--TYAPPECKHLEKRAVQGLCLMLNLTLQK 702  
QY 822 FSVTNVIRLKPSSVNVSPLEQAL--SQHGAHGNLLAVLAKYIYHKHPALPPLATQ 878  
DB 703 ENLPMDLRESGLAIYSPLEBLOGINPRTKKANVMI--ARYLYHGNN--PELAPES 759  
QY 879 LKRLATVAPMS-----VYACIGNDAA--IRDAFLTLQSKI 913  
DB 760 AKLICICISNSNIQYKMGVDFTHDQSVQKLMAGFEVCELDYDTEBFYVABEGSELEKEL 819  
QY 914 EDMR--IKVMILEFLTAVETOPGLILEP--LNLVYVDSGDSGKESISLGMW-----SCLHAY 966

DB 820 AAIHRETRHILNLLITLSERNPNLALYLLGFEIKKEPISITNLODPGVLGCPRTCLHAI 879  
QY 967 LELIDSOQOOR-----YWCPLRLHRAAIAFLHAL--WODRRDSAMLVTRTKPFWEH 1016  
DB 880 LNIIEKTEGSDGVAVREY--POLADLCYQVYIQLCAGSDTSGPTRYKRLTSODF--- 933  
QY 1017 LTPSLFGLTSP---PSETEPSILFETCALIKKIICELIYVVKSGSDLOSLKDTLKPSI- 1072  
DB 934 ---LFSGLQHLPSPSNHEHISMLSOWSMKTSIBLRV---SLNRQSHTRQLHL 986  
QY 1073 ---EKRFAYWSGVKSLAVHVAETEGSSCTSLIETOMLVANPMLIITATADIMHL 1127  
DB 987 LDDMPKVPSPDEGGM-----EDENRVSYSGFLHEDTATKVRKILSI----- 1028  
QY 1128 TDSV-----VRQFLVDYDQTKALLVPAVNCRL--LGSNCTLLILLRQMKREL 1179  
DB 1029 LDSIDFSGEIPQLQFPDQAQIBQVY--ANCEHKNLQOQVYCNVLL----- 1075  
QY 1180 SVDBILGFLTEILEGVLQADQOLMEKTKYAKVSAFIT---VLOMKKKVSDIPQYSQV 1235  
DB 1076 --HVLVAEVNALQGMALIGRPLMBEISTIIQVVGRRNLQCLHAKRBALESWRQV 1133  
QY 1236 LNVCEILOEVIALFDQTRHSLAGSATEDKDSMETDDCSSRRHDDQGVCVGLHLAK 1295  
DB 1134 EILITACPOELIOAEDR---OLIRDLQDVHDKVLD-----EAAQ 1172  
QY 1286 ELCEVDEBDSWLOVTRRLPLPLTLTLTLEVSLAMKQHLHTEATMILLTLAPQOAT 1355  
DB 1173 EL-----MPVAAVAFPL-----TAHLSQA-----VREBOROP 1200  
QY 1356 AVAGAGITQISICPLLSVYOLSTNGTAOTPSASR-----KSLD--APS 1396  
DB 1201 LVSGRGEQO-----YAFMLDSSLTSSPAESRPVGRASIGDSSMIHILKKLDPILKT 1253  
QY 1397 WPGVYRUSMSIMEQLKTLRYNPLPEALDVGVOERTLQCLNAVRYOS----- 1446  
DB 1254 GGGFORVYTHLYGSLLYLQIARQPBDEBDTBAAKTWERELTAPEDVPSKLORENMAII 1313  
QY 1447 -----LACLEAD-HWYGFILQS-----NMMEWHPLPOLMEDIQNLGYL 1488  
DB 1314 ESYGAALMEVYCRACDHEIGRMALALDRIVSVDKOHWMLYLS-----NSGL 1365  
QY 1489 COACTSLHSKMLQHYLONKNGDGLPSAVQVORPSAASAPSSKOPADTEASEQ 1548  
DB 1366 KVLVDSLVDDRTQSL-----TPQPLKALTYTESKMAFLTRVAREQ 1410  
QY 1549 Q 1549  
DB 1411 Q 1411

RESULT 24  
OGBLR9 PRELIMINARY, PRT, 1641 AA.  
ID OGBLR9;  
AC OGBLR9;  
DT 01-OCT-2004 (TREMREL. 28, Created)  
DT 01-OCT-2004 (TREMREL. 28, Last sequence update)  
DT 01-OCT-2004 (TREMREL. 28, Last annotation update)  
DT Similar to CA3071|CANUP188 Candida albicans CanUP188 Nucleoporin.  
GN ORFNames=DEHA0F122109;  
OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycetina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
NCBI\_TaxID=4959;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CBS767;  
RG GENOLABURES;  
RA Dujon B., Sherman D., Fischer G., Durrene P., Casaregola S.,  
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,  
RA Goffard N., Frangeul L., Aigle M., Anichoud V., Babour A., Barbe V.,  
RA Barney S., Blanchin S., Beckerich J.M., Beyne B., Bleykasten C.,  
RA Bolserne A., Boyer J., Cattolico L., Confantolero F., de Darvar A.,



DT 05-JUL-2004 (Tremblrel. 27, last sequence update)  
 DT 05-JUL-2004 (Tremblrel. 27, last annotation update)  
 DE AER284WP.  
 GN Name=AER284W;  
 OS Ashbya gossypii (yeast) (Eremothecium gossypii).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetaceae; Saccharomycetaceae; Eremothecium.  
 NC NCBI\_TaxID=33169;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 10895;  
 RX PubMed=15001715;  
 RA Dietrich F.S., Voegelé S., Brachat S., Larch A., Gater K., Steiner S.,  
 Mohr C., Pohlmann R., Ueidi P., Choi S., Wing R.A., Flavler A.,  
 Gaffney T.D., Philippen P.;  
 RA "The Ashbya gossypii genome as a tool for mapping the ancient  
 RT Saccharomyces cerevisiae genome."  
 RL Science 304:304-307(2004).  
 DR EMBL; AER16897; AAS52965.1;  
 DR ACD; AER284W;  
 DR InterPro; IPR002129; Pyridoxal dec.  
 DR PROSITE; PS00392; DDC\_GAD\_HDC\_YDC; UNKNOWN\_1.  
 DR SEQUENCE 1501 AA; 169463 MW; ACCD26E3015E829C CRC64;  
 Query Match 2.9%; Score 258.5; DB 2; Length 1501;  
 Best Local Similarity 19.8%; Pred. NO. 1.1e-07;  
 Matches 317; Conservative 220; Mismatches 561; Indels 507; Gaps 74;  
 QY 332 LLRHTLN-----ERTSSV-----VRKIGT-----AIQLN 357  
 DB 174 LRRPSNNVILYWSFVSVKAYLLEETSEVDFKXVFGTTTIDDLINFPAAABSLH 233  
 QY 358 VFQYTRLLQSLASGNDCTTSTACVCYGLSFLVLS-----ELHTGNQODIID 409  
 DB 234 VLDGICRLHKTLSDEG-----LYSAIMQSFLLISLNFISLNDETHAIXLVLTGMPELDIE 289  
 QY 410 TACEVLADPSLPELFWGTEPTSGLLIIDSVCGMPPHLSPLQLLRALVSGSKTRAKKY 469  
 DB 290 S---PLTSPLEKUL-----SILPAKLPLHESLIPMINL-----TSIHPEF 328  
 QY 470 SFLDMSFYNELKXKHPDV-----ISHEDGLWRROTPKLY-----PLGQTNLRI 517  
 DB 329 AHFKEKELSTYAKKANDIDYLDVDDN--RVESTDILVAKNELVYPRPEFEDNVL 386  
 QY 518 P--OCTVQVM--LDRAIVRWEXSYSSWTLF-----TCEIMLHVSTADVLQHC 566  
 DB 387 PIAKOTKQIIPSATGDEDVVV--FLYKSGMALIGLILQNICETVRDSCAPPTAARQYL 445  
 QY 567 QRVKPIIDLVKIVISTDLSIADCLPITRIMLQRLTVI----- 608  
 DB 446 --LVSIIDLISIVADTPL-----ERSTEIIQHSIGYVTEDDIISLIRTEBOALH 495  
 QY 609 SPPVIVASCNCLFVLAARNPAKVTDTLRHTGFLPFVAPVSSLSQMSAEGNAGGYG 668  
 DB 496 SRDMEVLVPLCLKLFVSMVTNPHIYKSH-----SKSDLDIRNG--KAGMT 540  
 QY 669 NLNANSPOQEGYGVTAFLRLITLVKQL-----GSTQGLVPCVMFVLKEMLPBY 722  
 DB 541 TLIGALELPNGEYLFITAFIKLADLVSESLHETAFPERMKRELLGRIVKOLLHYESY 600  
 QY 723 HKWRNSHGVRQIQGLIELIHAILNLCHETDHSHTSPSQFLCICGLATTEAGQYVI 782  
 DB 601 QVMKTSINQREIQLLSTLFTKLIYSYGIDODSKPGDKI-----TRVLADAASFIV 654  
 QY 783 NIMGIGVDTIDVMAAQRSDGEGQGLIKTKYKLAFTYNNVIRLKPPSNVSP-- 840  
 DB 655 TVF--IGTSDPD-----IRAVNSLVSV-----LVSPPT 680  
 QY 841 LEQALSQHG-----AHGNMLIAY-----LAKYIYK-----H 867  
 DB 661 LEAIIANGVNGAQYNNELLQKSFAPGNMLISVRLGLKLPPTLEKSIYKATDLDVAYI 740  
 QY 868 DPALRLAIQLLKLATVAPM-----SVYACIGN--DAAIRDAFELTRLOSKIEDRKIKYM 921

DB 741 KPSLKVHTIKFTYLVLR--APWFAERPSLLALHGNHAKSLLELTFDLSPIODRYVLKN 799  
 QY 922 ILFETLVAVR--TQGLIEFL--NLKVQSGD-----GSKEPSLGMSCLHAYELDS 972  
 DB 800 IYSPFSSIMEKQKQMSILFTGIVTFDDKHENKEVSK-----SILTIILKR 848  
 QY 973 QOQRYWCPPLIHR--AAIAFLHLMODRDSA-----MLVLRPKFWENLTPPLF 1022  
 DB 849 NALNDRPBDLGSHLDALISYARNSWTSARVYAADPITALVYKLE--ABEPVNDTLK 907  
 QY 1023 GLTSPSETSR-----PSILETCALIMKICLEIYVVGKSDQSLKDTLKR--PSIEKR 1075  
 DB 908 SIEQVASWASKYRLISRAICAL-----YLFTSGTSPFLDLGRUDLATIYVP 958  
 QY 1076 FAYMSGYKSLAVHVALETGSSCSLSLEYO--MLVSMRLLIATTHADIMHLDSDV 1132  
 DB 959 LFEIDGVNKDILHNLAVN-----FEKQMPGSLSTFSSTLVSDNSPQTIIYDPL 1010  
 QY 1133 RROFLVDVLDGTKALL-----LVPAVNCILRSGMKCTLLILLKQMKRELSGV 1181  
 DB 1011 MDQIF--SGDEAFGSESHTGAREQVVASIN--QPTVQISAASGALLTS-- 1060  
 QY 1182 DELIGPLTEILEGVLDQDLMEKTKAKVFSAFIV--LQMKMKVSDIPQYSQVLVN 1237  
 DB 1061 -----PIKTPVPLQDTFVDIVLHLKVNIEKSVSPFVDIYLE 1100  
 QY 1238 VCETIOEIVIALPDQTRISLALGSAETEDKSMETDDCSRRHRDQDQCVGLHIAEL 1297  
 DB 1101 RIF--LCFVLYLSFMKTSKVV--EKNIVE-----ITSHMKIL 1135  
 QY 1298 CEVEDGQSWLOVTR--LPILPTLLTLEVSLRMKNLHFEAT----- 1340  
 DB 1136 TSSEIDPLGNIASSKSAVYRPLRSILIALSL--VATGHFIEVVELLEPERIFC 1192  
 QY 1341 --LHLLTARTQOQATVAG--AGITQSI--CEPLSYQSLSTNGTAQTPSARKS 1391  
 DB 1193 KGVNLIISLISEINTTSSCGNNVITSVTXIQDLILLISLF-----TXIKN 1240  
 QY 1332 LDAPSPGVYRLSNLSMEQLKTLRYNPLPALDFVGHQERTLOCLNAVTVQSLACLE 1451  
 DB 1241 LQPFK-----SEFLM-----IATSLNEVDITIKALNLY 1268  
 QY 1452 EADHTVGTIQLSNMKEMHFLPOLMRDIQVNLGYLCOACTSLHSHKMLQHYLQNNKG 1511  
 DB 1269 SSSHL-----FRVNEEPIIADITLFLCEFSVDEVAEKLII----- 1305  
 QY 1512 DGLPSAVARVQRPASASAPSSSKQPADTEASEQALHTV--QYGLIKLISKTLAL-- 1569  
 DB 1306 NGLPSVLE-----SPISIMIOGGGKKEVQRLHSIWENGLSIIQLLSQFG 1354  
 QY 1628 LDKKK-----BPLQAV-----GLSTQABSTRILKSLMTMENCFLYLSQMR 1672  
 DB 1407 LEYQCYLNSNAKTKVIDDVVIEIIPGLDLELEKELIYS--FT-----HLTHP--K 1456  
 QY 1673 YLRDPAVHR--DKQRMKE-----LSSELSTLSSIRYP 1706  
 DB 1457 YLNSKIVPTIEQQLQDLDETRAAFYKNTTBETIRKQDSLSESP 1501  
 RESULT 26  
 AAS52965  
 ID AAS52965 PRELIMINARY; PRT; 1501 AA.  
 AC AAS52965;  
 DT 23-APR-2004 (Tremblrel. 27, Created)  
 DT 23-APR-2004 (Tremblrel. 27, Last sequence update)  
 DT 23-APR-2004 (Tremblrel. 27, Last annotation update)  
 DE AER284WP.  
 GN AER284W.

OS Ashbya gossypii (Yeast) (Eremothecium gossypii).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Eremothecium.  
 ON NCBI\_TaxId=33169;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 10895;  
 RX PubMed=15001715;  
 RA Dietrich F.S., Voegelé S., Brachat S., Lerch A., Gates K., Steiner S.,  
 Mohr C., Pohlmann R., Luedi P., Choi S., Wing R.A., Flavier A.,  
 RA Gaffney T.D., Philippen P.,  
 RT "The Ashbya gossypii genome as a tool for mapping the ancient  
 RT Saccharomyces cerevisiae genome."  
 RL Science 304:304-307(2004).  
 DR EMBL; AE016897; AAS52965.1;  
 SO SEQUENCE 1501 AA; 169463 MW; ACCD26E3015E829C CRC64;

Query Match 2.9%; Score 258.5; DB 2; Length 1501;  
 Best Local Similarity 19.8%; Pred. No. 1.1e-07;  
 Matches 317; Conservative 220; Mismatches 561; Indels 507; Gaps 74;

332 LIRHTLNP-----BETSV-----VRKIGGT-----AIQAN 357  
 174 LIRPSNPVILYWSVTLVKAYLLESTBNVDFIKVFGTTTIDDLINTFAARASLH 233  
 358 VFQYLRLLQSLASGNDCTTSACMCYGLLSFVLTSL-----ELHTLGNODDID 409  
 224 VLDGICRLHKTLSDEG---LYSALMCSFLILSLINSLANDETTHAKIVLGMFDDLE 289  
 410 TACEVLADBSLELFWGTEPTSGLLIILDSVCGMFHLLSPILQILRALVSGKSTAKVY 469  
 230 S---PLTSELEKCL-----SILRAKCLPLHESLIPMINTL-----TSIHPER 328  
 470 SFLLDKASFYNELYKHKPHDV---ISHEDGLWBRQTPKLLY-----PLGGQTNRI 517  
 329 AHFEKEKELSTYAKQKANDIDYDSDDN---RVSELDILVAKNELVYPERFEEDVNL 386  
 518 P---GGTVGVM---LDDRATLVWEXYSYSSWTLF-----TCSEIMTLHVSSTADVIOHC 566  
 387 PIAKOTKQGIIPASGDEDEVVY-FLYKYSGMALIGRIILQNICETARDSCAPTAALROYL 445  
 567 QVRKRIILVHKVISTDSLADCLLPITRIMLOLRLTVI----- 608  
 446 -LVSLIDLISIVADTLP-----ERSTEIILOHISGYTEDDIISLIRTEQALH 495  
 609 SEPVDVIASCVCITVLAARNPAKYWTDLRHTGFLPVFAHPVSSLSOMLSAEGMAGGVG 668  
 496 SRDKEVLVPCLKLFISMWTNPHIYWSH-----SRSDLDIRHG-KAGMVT 540  
 669 NILMNSEQPQGEYGVITIAFLRLITLVKQGL-----GSTSQGLVPCVMFVLKEMLPSTY 722  
 541 TLIGALELPNGEYLFITIAFIKLDELIVSESLHETAFPERMRKELGRIVKQLHIVSY 600  
 723 HKMRNHSQVBEQIGCLIELHALNLNCHETDLSSHTSPLOFICISLAYTEGQVVI 782  
 601 QYWKYSINQRFEGALLTSFTKLIYSYGGIDQSKRGDKI-----TRVLDAASRIV 654  
 783 NIMGIGVDTIDMVMAOPRSDGAEQGGQQLIKTVKLAFTVNNVIRLKPSNVSP-- 840  
 655 TVP--IGTDSPD-----IRAVNSLVSV-----LVSPFTT 680  
 841 LEQALSQHG-----AHGNNLIAY-----LAKYIYHK-----H 867  
 681 LEAISLNLNGVGAQYNELLQKSFAGNMILISVGLLKPPTTLSEKISYKATDVDAIY 740  
 868 DPALPRALIQLLKLATVAPM-----SYAACGN--DAALIRDAFTLRQSKIDEMRIYVM 921  
 741 KPSLKVHVIKLETVLVR-APWFAERPSSLAHGHAASLLELTFDSSSPIDRVLYKN 799  
 922 ILBEETVAVE--TOPGLIELFL--NLEVNDGSD-----GSKRFSLGMSCLHAYLELDS 972  
 800 IYSPFSISMEGKQGMSTLFLTGNIVTVDSDSKHEKVGSK-----SILTILKR 848

QY 973 QGQDRWCWPILHR---AALAFHLAMODRPSA-----MLVLRTRKFWENTLSPLE 1022  
 DB 849 NALNLDKRPEDJGSHLLDIAISTAFNSWTSARVADKOPTALVRLKLE-AFEPNVNTLK 907  
 QY 1023 GTLSPPSETSE-----PSILETCALIMKIICTEIIYVVKSGISDQSLDYLTK--FSIERK 1075  
 DB 908 SIEQVYASMAKSKRLISRIALICAL-----YLTSGTTSPIFDLGRDIAITVNP 958  
 QY 1076 FAYWGYVKSALVAVALEFGSSCTSLBYQ--MLVSMRMLLIATTAADIMHLDTSVV 1132  
 DB 959 IPELDGKMDLNTNLAIV-----FEKOWEGLSLSTFSISTLVDSNSFOTIITVDLPL 1010  
 QY 1133 RQGLFDLVDGTKALL-----LVPSVNCILRGSMKCTLLILIRQWKEIGSV 1181  
 DB 1011 MDQYF-----SGDEANFGSESHGLRQGVASINL-----QVTTQISAAKSGALLTS- 1060  
 QY 1182 DELIPLTEILEGLVADQOLMEKTKAVFSAFITV-----LOMKEMKVSIDIPOYSOLVYN 1237  
 DB 1061 -----FIKKTVPVLPQTFVDIVLHLKVNIEKVSSEPFVDIYLE 1100  
 QY 1238 VERTQAEVIALFDOTRSLALGATKEDKSDMETDCCSRSHRQRDQVCVGLHLAKEL 1297  
 DB 1101 RIR-LCFYLYVSFMKTSKQVNT-----EKNLVE-----ITSHMKIL 1135  
 QY 1298 CEVDEEDGDSMLQVTRR-----LPILPTLLTLEVSLSMKONLHPTAT----- 1340  
 DB 1136 TSSEIDPFGNIASSKSKAYRPLRSILIALSL--VATGTHFELVSEVLELFEPRIFC 1192  
 QY 1341 -LHLLTLTARTQCATVAG-----AGITQSI--CLPRLSVYOLSTNGTAOTPSASRKS 1391  
 DB 1193 KGVNLIIEILSEILSEINTTSSCCANNVSIETVTKIDILLLSLIF-----TKIKN 1240  
 QY 1392 LDAPSPGQVYVLSMSIMEQLKTIKRYNPLPALDPVGHQSRITLOCLANAVTQSLACLE 1451  
 DB 1241 LQPER-----SEKLM-----IATSLAEVDTIKALNLY 1268  
 QY 1452 BADHTVGIQLQSNFMKEMPHLPLQMRDIOVNLGYLCOACTSLHSRKLQHYLQNRNG 1511  
 DB 1269 SSSHL-----FRVNEBPIADITILFICEPSSVDBVAEKLI----- 1305  
 QY 1512 DGLPSAVQRYQRPASAASAPSSSKOPADTBASBOQALHTV-OYGLKLISLTAAI- 1569  
 DB 1306 NGLPSVYLE-----SPISIMIQGGIRKEVOPRLHSIMNSGILSIILQLSQFG 1354  
 QY 1570 RHFTPDVQQLIDQSLDAEYVFLFALSTPTPTDSVAPRFGTLLATVNVANL--NMIGE 1627  
 DB 1355 AQVLPBCLFTYYSKQSLSTIFSWS-----DNSLALSNSIITQDSQIIMQKMFSA 1406  
 QY 1628 LDKKK-----EPLTOAV-----GLSTQAGETRLKSLMFTWENCFYLLISQMR 1672  
 DB 1407 LEYQKIVNSNAKTKVIDDVVIEIIPGLDELERKELYS--FT-----HLTHP--K 1456  
 QY 1673 YLRDPAYHPR-DKQRMQOE-----LSSEISTLSSLSRYF 1706  
 DB 1457 YLNSRIVPTLLEBRQLQDEBTRAFAVKNTBEIRKLODSLSPSEF 1501

RESULT 27  
 Q6FP54 ID Q6FP54 PRELIMINARY; PRT; 1670 AA.  
 AC Q6FP54;  
 DT 05-JUL-2004 (Tremblrel. 27, Created)  
 DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
 DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
 DE Similar to ep|P52593 Saccharomyces cerevisiae YML103c NUP188 nuclear  
 DE pore protein.  
 GN ORFNames=CA6J0J0634g;  
 OS Candida glabrata (Yeast) (Torulopsis glabrata).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 ON NCBI\_TaxId=5478;  
 RN [1]  
 RP SEQUENCE FROM N.A.



Db 1539 SNOGNTLV-----SFWDSMNSVSTALVNETKQITL-----LQKML 1570

Oy 1633 EPLTQAVGLSTQ-----AEGTRTLKSILMTFWEICFYLLISQAMRYLRDPAVHR----- 1680

Dd 1575 EAINKVNSVSTSRNNSVTQNPNDNVKQJPGVLVSFDYQKOLNSALKKL-----THPKYLSNR 1630

Oy 1683 -----DKOMXKEIJS--SELSTLSS 1701

Dd 1631 IVASTTDGEHMEIEISRRELSTLIISN 1657

RESULT 28  
N184\_SCHPO STANDARD; PRT; 1564 AA.

ID N184\_SCHPO  
AC O9P7M8; Q9YG64;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Nucleoporin nup184. (Nuclear pore protein nup184).  
GN Name=nup184; ORFNames=SPAP27G11.10c;  
OS Schizosaccharomyces pombe (Fission Yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_taxid=4896;  
RX (1)  
RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.  
RC STRAIN=972;  
RC MEDLINE=993118821; PubMed=1038805;  
RA Whalen W.A., Yoon J.H., Shen R., Dhar R.;  
RL "Regulation of mRNA export by nutritional status in fission yeast."  
RT Genetics 152:827-838(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RC MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouras J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Fellegue J.T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holtroyd S., Hornsby T., Howarth S., Huckle R.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., Mclean J.,  
RA Mooney P., Moule S., Mungall K., Murphy J., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbinowitsch B.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Riesger M., Schaefter M., Mueller-Thuer S.,  
RA Gabal C., Fuchs M., Fritz C., Holzer E., Moesl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehnach H., Reimhart R., Pohl T.M.,  
RA Egger P., Zimmermann W., Wedler H., Wandut R., Furnelle B.,  
RA Goffeau A., Cadieu E., Dreno S., Gloux S., Lelaure V., Motier S.,  
RA Galibert F., Aves S.U., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Galliardin C., Tallada V.A., Garçon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shipakovski G.V., Ussery D., Bartrell B.G., Nurse P.;  
RT "The genome sequence of *Schizosaccharomyces pombe*,"  
RL Nature 415:871-880(2002).  
CC -!- FUNCTION: Interacts with pom152 in the core structure of the  
CC nuclear pore complex (NPC). Involved in the export of mRNA.  
CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.  
CC -!- CAUTION: Ref.1 sequence differs from that shown from position 1560  
CC onward and is longer (1628 AA) due to a frameshift.  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (see <http://www.fdb-sib.ch/announce/> or send an email to [license@fdb-sib.ch](mailto:license@fdb-sib.ch)).

DR EMBL/AF055035; AAD43830.1; ALT\_FRAM.  
DR EMBL/AL157917; CAB7631.1; -.  
DR GeneDB SProbe; SPAP2761.10c; -.  
KM Nuclear protein; Transport. S -> P (in Ref. 1).  
FT COMPACT 1219 1219  
SQ SEQUENCE 1564 AA; 16962 KM; 2483AEF38C30B7 CRC64;

Query Match 2.6%; Score 233.5; DB 1; Length 1564;  
Best local similarity 19.4%; Pred. No. 5.4e-06;  
Matches 332; Conservative 250; Mismatches 663; Indels 469; Gaps 74;

QY 34 SGIEMLNKMRRLLEG---LSTYKPR--SPSSAEKVANDVASPLKELGLRISKFL 86  
DB 43 SPVGTGLNSN--KITEGEBRATLSYDSHSISKSQLESVK-----KISDIT 86  
QY 87 GLDEESQVOLL---QCQYQEDYGRTRGSVKTVAQDEROSQALILKADYYERTG--IL 141  
DB 87 GYNEAQVAVYLVNQLNYELNTQYFSQULN--DSVLAQSFQR-----YIABIISCKVVL 137  
QY 142 RCVTHLITFYDERHPYRVEYADCVDRLKEIKELVSKYRQ----- 179  
DB 138 AFLTACTDASDKVN-----KMATRLIVSIYFQTAQSGENAGNSTPSPICRIT 186  
QY 180 -QFEELYTEAPWETHEGNMTBTRQVRWVQCLRQSNLEIIFLYAYAFKAPSDLLV 238  
DB 187 DYSMTSQAPASPLTNG--BEAHSQWFFHFNTQLDLRLTYL-----STYSLLV 236  
QY 239 LTKQFKEGGFSRRQTNHLYDE-----TMDPFVDRIGYSALILVGMIDIESLHKALD 293  
DB 237 CNSEMAISWPFCKMKCTRYLHDQEPMLHIDITGFSMCKEITNVAIIISNIFSLKQVLSF 296  
QY 224 RREIHQPAQDGLIQDMDCLMTLFGDIPIHAPVLLAMALRRTL--NPZ-----ETSSYV 346  
DB 237 KDNPSFPLSGNTTIIISLHMDITQLSDNSIGAAVSLTWGIALHLNSPNPILQINSSVY 356  
QY 347 -----RKIGGTALQLVFQVLFELTLOSLASGGNDCTTSPACGVGLSPVLT 394  
DB 357 SSKILQNPQNSFQALIIAALKQDPFLIRIIRIISL-----EDDPYIDGYSKMATLF 408  
QY 395 SLEHTGNOODIIDTACVGLADPSLEPLFWGTEPTSGLGIIIDSVCGMFFHLSPLLQL 454  
DB 409 SSASVYKFSDSMTLCYTLTFTKPOYQULF---ENNDSYTRRLNPARAFPRYSQFVL 465  
QY 455 LRALVSGKSTAKKYVSPFLDKSPFYNLYX--HKPHVISHBEGT---LMRRQTPCLATPL 509  
DB 466 LIIPFACLTGKQVSSBELHMTTFQSLDPSGFKAYEIIIEPVNTGNALIELOSLLHDSY 525  
QY 510 G---GQTNLRIPOGVAGQVW-LDDRXYLVKWEYSYSWTLFTCEIEMLHVSTADVIOH 565  
DB 526 GFPEFPAERSLPGTGTGRIVSVDTYPPVVMMDLNSLMEAVGISLN---YIIRNGELINSH 582  
QY 566 CQRKPKIIDLVHNVISTDLSIDCLPISR-----IYNLDRLLTVVISPPDVIVAS 617  
DB 583 KSFVLTVLSSVPLFQTDVSGACELVHLASBGLDELPFINVICDLDYFASLSTVEDAD 642  
QY 618 ---CVNCLTVL-----AARNPAKWTDLRHGFLPFAVHPVSSQMSIOMISGMNAGYG 668  
DB 643 YQICVSSLRRLREFTFAATD--YV-----AYTRSLVLCVGSSEKGIS--LE 684  
QY 669 NLMANSEQPOGEGVTIAPRLI-----TTLVKQLDSQSGLVPCVMFVLKEML 719  
DB 685 DIVFDEHSINGVDFTLAEPFLYEIILDNICISTSVAPDPFSIRLKTDPVKQAMFLCEVF 744  
QY 720 PSYHKRRYNSHGVRBIGGLIIEHLALINLCHETDLSHSST----- 761  
DB 745 ANYLDMKXARITIQQYQIGRRFASLLTYKLNLNTPGIIIEYRNPRTYVKKTLPLRBSHIYVQ 804  
QY 762 -----PSLQFLCISLAVTAGQGVINIMIGIVDTIDMWMAAPRSDGAEQ 808



Db 805 RFLVQDSNRNRYLHPLLSVMDLNLITDITPSTI-----SSPRAKAAK-- 846  
 Qy 809 GGGGLIKTVKLAFTVNNVI-----RLKPPSNVSPLEQALSGHAGNNLAVLAKY 862  
 Db 847 ---MMLIS---SFCAMKTLICRGFLNLKP-----SELRELPERSPDLPN----- 886  
 Qy 863 IYHGHDPALPRL-----AIGLLKRLAVAPM-----SVVACLGDAADAIADAFRLQS 911  
 Db 887 ---CLPRLCCIAPIQLLSAL-ILAPMSETPSLLYMINSTDIVGVCQILTN 938  
 Qy 912 KIEDMRKIMLEPFLTVAVE--TOPGLIELFLNLEVKDSDSKERSLG--MNSCLHAV--- 966  
 Db 939 PIQSTNIEGSAWKLSSIMKQOQGLAVLFS-----GKKPPLDRMKSLLNNVVO 989  
 Qy 967 ---LELIDSOQODRYWCPPLHRAIAFLHLMQDRDSDAMVLRTKPKFMENTIS-- 1019  
 Db 990 LTSKSLISIAEKRLDSFINDILSQVPFEFFPLSRNFWTSLGTLQOENAFMNRIVDAI 1049  
 Qy 1020 --PLFGTLPSPSEPSSEPILETCLIMKICLEY-----VYKSLQSLKDT 1066  
 Db 1050 KLPLTVKLDGLSVAQADLYLLAAATRTITAIQLHMSKLKNSSKKIIIDPLKDSMKDL 1109  
 Qy 1067 LK-KFSIEKRPAYNGVYSIA-----VHVAE----- 1092  
 Db 1110 VQHAFTIT--AYDSNHNALTRAFKHNGDLHIDLNTGLFLRYGDNYFYNNKLAKN 1166  
 Qy 1093 ---TEGSSC-----TSLELYQ--MLVSAVRMLLIITTHADIMHLDTSVVRQL 1136  
 Db 1167 MLNLTEDTSPFKISWMSANENSLIDQAALLRSW-SIFICAFVEFKEDATLSTLEKI 1225  
 Qy 1137 F-----LDY-----LDGTQALLLVPSVNCAL--GSMKCTLLIILRQK-- 1175  
 Db 1226 MKWVLSIAEDTIDVNVVQELSAERPAALVFRISQOTLAIPISENVEKHLQSLITLWAI 1285  
 Qy 1176 --RELSGVDIELG-----PLTEILEGYLQADQOLM--EKTAKYPSAFIT-VLOKEMK 1224  
 Db 1286 TTYTFSITEDSNGEAAVYRPLHAYNTL--NRLSEKENTLSVGFVSLDLCRHK 1342  
 Qy 1225 VSDIPQYSQVLANVCETIQOEVIATFDQTHSLAGSATDKDSMETDSCSRHRDQD 1284  
 Db 1343 LSQI--PEKAVINPIETEVYGDIVL-----NSLHK----- 1370  
 Qy 1285 GVCVGLHLAKEL--CEVDEDEGDSWLOVTRRLPIPLTLTIFSVLRKQNLHFEATLH 1342  
 Db 1371 --CYNSSHILRGLOSLYSYINDSP--SYDNCL-----RLFSWSHSLVDGQYFADALIS 1422  
 Qy 1343 LLTLTARQOGATAVAGITQSIGCLPLLSVYQLSTNGAOTPSASRKSLLAPSPGYR 1402  
 Db 1423 FLTICSSPAGAEQIVNNGFFYSIMESPST-ALSTGGLGDS--SIQYKIN---IR 1474  
 Qy 1403 LMSMIEOGLKTLRYNPLPEALDPVGHQERTLOCLINAVRTQ-----SLAC 1449  
 Db 1475 GILPLFLPIVIFKLRINMNDREPV-----LLAFPOQVALLMWCPSPSISLAS 1524  
 Qy 1450 LEBADHTVGTILQLSNPKEMHFLPOLMRDQV 1483  
 Db 1525 IDES---FMIVLPLDLQCFN---PALQBIRL 1551  
 RESULT 29  
 Q6FP99 PRELIMINARY, PRT, 1710 AA.  
 AC Q6FP99,  
 DT 05-JUL-2004 (TREMblrel, 27, Created)  
 DT 05-JUL-2004 (TREMblrel, 27, Last sequence update)  
 DT 05-JUL-2004 (TREMblrel, 27, Last annotation update)  
 DB Similar to sp|P47054 Saccharomyces cerevisiae J01039c NUP192.  
 GN ORFNames=CAGL0J00759g;  
 OS Candida glabrata (Yeast) ('Torulopsis glabrata').  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; microsporitic Saccharomycetales; Candida.  
 CX NCBI\_TextID=5478;  
 RV [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=C86138;  
 RG GENOLAYURES;  
 RA Dujon B., Sherman D., Fischer G., Durrens P., Caaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla B.,  
 RA Goffard N., Prangeul I., Aigle M., Anthouard V., Babour A., Barbe V.,  
 RA Barnay S., Blanchin S., Beckereich J.M., Beyne E., Bleykasten C.,  
 RA Boismarie A., Boyer J., Cattolico L., Confarotier F., de Daruvar A.,  
 RA Desorme L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 RA Hantaye F., Hennequin C., Jauniaux N., Joyet P., Kachout R.,  
 RA Kerrest A., Kozul R., Lemaire M., Lesur I., Ma L., Muller H.,  
 RA Nicaud J.M., Nikolaki M., Ozcas S., Ozler-Kalogeropoulos O.,  
 RA Pellenc S., Porter S., Richard G.F., Straub M.L., Suleau A.,  
 RA Swenne D., Tekala F., Mesolowski-Louyel M., Westhof B., Wirth B.,  
 RA Zenhou-Meyer M., Zivanovic I., Bolotin-Pukhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,  
 RA Wincker P., Souciet J.L.,  
 RT "Genome evolution in yeasts".  
 RL Nature 430:35-44(2004).  
 DR EMBL: CR380956; CAG60692.1;  
 SQ SEQUENCE 1710 AA; 195270 MW; 4915517790EC73 CRC64;  
 Query Match 2.6%; Score 232; DB 2; Length 1710;  
 Best Local Similarity 18.1%; Pred.No. 7.7e-06;  
 Matches 325; Conservative 281; Mismatches 602; Indels 590; Gaps 84;  
 Qy 17 SREMTILGSAARELSQIEAEINRRLLEGLSTYKPPSPSAEKVANKOVASPLK 76  
 Db 21 SSELFTQL--PDKEIKLKPCK-NTSSRNLEKLDQKTDGSI--FKLNQDFTSI- 73  
 Qy 77 ELGRISKFTGLDEOSVOLQCYLOEDYRGTQDSVKTVLQDEROS--QALIKIADY 132  
 Db 74 ---ISLDELNLDELAVNL-----VLNSPESQIYASVNVNGKISFLRQYIQLIYIS 126  
 Qy 133 ---YY-----BEKTCILRCYL-----HLTYFQDERHPVREYADCYDK 168  
 Db 127 ICNADWSPYKKFPAEATNTTILSKVLENFDIISKLSIDIKQBN--RQQLDMNMP 183  
 Qy 169 LEKELVSKYRQGFELKTEAPT--METHGNLMTERRQVSRMAYVQCLRESMLEIIFLYA 227  
 Db 184 LITGIIHKRDPFESKEVDLLSQVAYGLSKRELINTQISS--VDLIIEHSLMETNDPFPV 241  
 Qy 228 YF-----EMAPSDLVLTQMP-----KEGCF----- 248  
 Db 242 YPLPLTLNLFENLNTFSDVIDTFLGHKKTNELITKQEDVYKPKMVLIMFTFLTGMC 301  
 Qy 249 ---GSRQTNRLVDETMDP--FYDRIGYFSALIL-----VEGMDIESLHK- 288  
 Db 302 KEDPQNRVTKANFSDIDDPWTIAVELGATQGLIIVADLSRIDESTAVQYDFRALER 361  
 Qy 289 ---CALDREBLHOPADGLICQDMCLMTFGDIPIHRAVLLAMALLRHTLNP-- 339  
 Db 362 HLPFTIPLQVDEVNITCKYQMSLTERSDSTKI--NGPHE-----SFSRNTYRPPS 411  
 Qy 340 ---BETSVVRKXIGGTALQNL-----VF-----Q 360  
 Db 412 MISPNNTYSKDRFTNGLASTISIMDRLEGEVFTSTDMKKEISKFRENISEMPEMFLB 471  
 Qy 361 YLTRLLQSLAGGNDCTTSTACMCVYGLISFVLISL--ELHTL--GNQODIITPACVTL 415  
 Db 472 NIHHIQTFFIS--DC-----AFILITLKOABDSLSLGSDDLDDIATK-- 513  
 Qy 416 AD-----PSLPELFWGTEPTSGLGIT--LDSVCGMFPPLISPLQL--LAL 458  
 Db 514 ADLERFVSVFYFSKRPYVAALFWEDEKSAAYGFEWASAKCN--DNLMKSCFYMLMGL 571  
 Qy 459 VSGKSTACKVYSPIDKMSFYNELVYKRPDVISHEDGTLRQRPKLYPLGGGTNLARP 518  
 Db 572 SSGKKNALNVTYTL-----GDSNFTSMK-----IADTIGISYSKILN 609  
 Qy 519 OCTGVQWLLD-----RAYLVKWEYSYSMTLFTCEIMLHLVAVSTADVIGHQCRVKPII 573  
 Db 610 LKALQAKGQKRGELNTTLALBGLNBERITVPLS-SLITLISVADQMBPBDKLM--- 665



```

QY 574 DLVHKYISFDLSIADCLPITSRTYMLQRLTIVISPPVDIVASCVNCLTVLA---ARNP 630
DB 666 -FSHLVADVIFELTKETPL-----VACENNTISTIPILIESDR 703
QY 631 AKYWTDLRHTGFLPVAHVPSSLSQMSIAGNAGVGNLWNSBOPOGEYVTIAFLRL 690
DB 704 ASIMHKLDSST--LFGKVALSDTSQYISA-----FSSILTT-----YSDVAGFLAL 746
QY 691 ITTLVX-----GQGST-QSGVLPCVMVVKEMLPSTHKM--RRN 728
DB 747 IRLNLIQIESFDCNKEFLVRGHLQYPNOLGAPYKRSIMWPLEFKNMVEFVNSDKITDSFK 806
QY 729 SHGVEBOIGCLILELILHAILNLCHEPDLSSHTPSLOFICGLATBAQGVYINIMGIG 788
DB 807 RHELDOR--VLEIT-----LTCNSFDYS---VILNLSISVG 837
QY 789 VDTIDNV---MAAPRSDGAEQO---GQGLILITVXKLASVTNNVIRLKPNSNV--- 837
DB 838 ANLDMLVHGENPASYVDNAAPAMFNFLQENVTYTLGYSIASVGIDNSLAKPHQLQTL 897
QY 838 ---VSPLEQALSCHAHGNNLAVLAKT-----YHKM----- 867
DB 898 IRLSLSVLERTLDYKXTYEELTTIKKYQSGEFYKPKYTGHLGVSFTQVFEFLDPVT 957
QY 868 -----DPALFRLAIQOLKRLA---TVAPMSVYACIGN-----DAAAIRPA 904
DB 958 CGLVYIGLNDISIPISINILAKRLALPSTEDPPL--LGNKFLTPSIDESARIKHA 1014
QY 905 FLTRLOSKE---DMKIKWMLLEPLTVAE---TQGLLEPLFNLVNDGSDGSKPFSIG 958
DB 1015 FIDOLDCEPIENDNMLKVEILEFLDSDIDSISSKSVNHLGFGY---DNTLTFGN 1070
QY 959 M-----WSCHAVLELIDS---QOQDRY---NCPPLHRAIAPLHLMQRRSAMLV 1006
DB 1071 LSTFQSGHSLFHSYIRLILSALNLSSTYKIDPAPRLAALSQLEIMCKSQLTSDVLT 1130
QY 1007 LRTKPFMENTLSP--LFGTLPSPSTSEPSI---LETGALIMKICLLE-----I 1051
DB 1131 -----EHITNNEIKIKMOVDPQTSHTLMNGINLESTQSKARFIQSSAGAYLM 1182
QY 1052 YVYVAGSJDQSLKOTLKKEFSIE---KRAYMSGYKSLAVHVAETEGSSCTSLLEYQMLV 1108
DB 1183 FLRYTSMIQFLNVLVHTLSLNGTISRKAPYYSILINSIYSA----- 1225
QY 1109 SAMRMLLIATTHADIMHLDTSV---RRQLFLDV-LDGTKALLLVPAVNSCLRLGSMKCT 1165
DB 1226 ---RIFSLRLPRLHONIT-LKEXIVPENQQLKDLFVDDKVLKRSCTGNLYDFDEISSV 1281
QY 1166 LLLILRLQMKELGS---VDEILGPLTEI-----LEGVL 1196
DB 1282 LELHSKTVSVSQNSKEFMDKILSFEFVKQILTLFLVRKSFSGOQLAIHNNWQVQII 1341
QY 1197 QADQOLMKTAKYVSAFTVLOKEMKYSIDPOYSQVLNVCEFLQSEVIALPQTRHS 1256
DB 1342 VSDNGLGRKR---STPI--LEVPMVTPKISVYVDIDVKAEBELVSLMWFLEFIYQOS 1395
QY 1257 LALGSATEDKDS---METDCCSRRH 1279
DB 1396 NASSSGKFLDSKLANLFOVCIRGISPLSSIKLASDYVYANNYLAVLNKSGMAVATV 1455
QY 1280 RDOR---DGYC-----VLGLHAKELCEV-DEDDGSMLOVT----- 1311
DB 1456 KDLRLNBRLEVAICNDALYEGEAKIKITGLLLDLSVLQGNHNGNFIWMLMSTKLFH 1515
QY 1312 ---RULPILPILITLLEVSILAKKONH---FTBATLHLLTLARTOGATVAVAGITQSI 1366
DB 1516 IIRSLKATDLSLNSSTENVNENFLYELTAKATVPLIRIARTRBGAVALIESKVPDIL 1575
QY 1367 C-----LPLLSVYQLSTNGTACTPPASRSKL--DAPSWPGVYVLSLMBQLL 1412
DB 1576 AELNPLKIDPDLGELVFDETSNNITS---SLRRVNTTDEPLIIGKEATSVSIFELIIT 1630

```

```

RESULT 30
Q7SY48 ID Q7SY48 PRELIMINARY; PRT; 2159 AA.
AC Q7SY48;
DT 01-OCT-2003 (TREMblrel. 25, Created)
DT 01-OCT-2003 (TREMblrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Hypothetical protein zgc:63510.
GN Name=zgc:63510;
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_Taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Whole body;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Feingold B.A., Grouse L.H., Derge J.G.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedtin T.B., Toshituki S., Carninci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Adamson R.D., Mullahy S.J.,
RA Boeck S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richardson S., Wozley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Heltón B., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shervchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalske U., Smalins D.B., Schnerch A., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Whole body;
RA Klausner R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055128; AAH55128.1; -
KW Hypothetical protein.
SQ
SEQUENCE 2159 AA; 242048 MW; F00DBA1D597E70B CRC64;

Query Match 2.4%; Score 218; DB 2; Length 2159;
Best Local Similarity 18.2%; Pred. No. 9.6e-05;
Matches 406; Conservative 300; Mismatches 653; Indels 876; Gaps 106;

QY 1 MIRSKITSVL-----SFCRSSRLMTLLGRLSALRELSQLIBALINKMRRLLEGSLY 53
DB 20 LLAGKEVAVSVLPDKDAASMDRSTFH---ALCGTGLBELMGIDAPRSEFQETLFSQSL 75
QY 54 YKPPSSAKRYKANDQVAPLKEGLRISKPLGDLBEGSVQLQCTLOEDRYRTORSYK 113
DB 76 -----TLRISVQSEVN---KKDKSISLFL----- 98
QY 114 TVLQDERQOALILKADYVEERTCILRCV-----LHLLTYFQDERHPRVVEYACVD 167
DB 99 -----TLSPYPLKPR-----LKCIEMLLRFHILYNQDILACVLPYHE--- 140
QY 168 KLEKEVSKTRQOPEELIYKTEAPTWETH-----GNLMTROVSRNVF--- 209
DB 141 -----TKVFAVVIQFLKIEDPTHKMHNGIQKGVPLARGLTLTHCYKDLGFMDFVC 193
QY 210 -----QCLRSQNL-----LEIIFLYAYFEMASBDLL-----VLTGM--FKEGFG 249
DB 194 SMTNSVKAYSBELTRDQNCQPLAVITSPFASSTIVSALDAVEKTIINSTIALGLPFDGLK 253
QY 250 SROTNRHLVDETMDPFVDIRIGVPSALILVGEAMDIESLHKCALDDBRELHQPADGLICOD 309

```

```

Db      254 S-----NLSDYTAATYM-IVCOMAVKVMBAQVDSL---SVQSRSGRTPQ--LIREG 302
Qy      310 MDCLML-----TRGDIPIHAPVLLAMLLHNTNPBSTSVARKIGTAQL 356
Db      303 LSCIIILLQNKQKVGIGKTYG-----YLCAVPLVSTLOSISTVHIDPLSLYLPH 355
Qy      357 NVFQYLRLLQSLASGGNDCTTSTFACVCVGLSLFVLSLELHTLGNQODIIDPCEVLA 416
Db      356 LIHSVMQNDQOQNEGLSDST-----GLLOSFPQNLQ--SSNIENIRAKLL 401
Qy      417 DPLSPEFWGTE-PTSGIG----- 434
Db      402 E---EYVVCNENLPSDDGISALNQRIOPTVRLPESRYPICALDMALENHYKNSDNEKLL 458
Qy      435 ---TILDSVCGMFPHL-----LSPILQLRALVSGSTAKKYSPFLDKSPFNELX--KHK 485
Db      459 HQFISLTLSCKGYOILPESETSLMISLHNPPLPSVANA-----VDYKEILINSBN 509
Qy      486 PHDIVSHEDGTLMR--RQTPKL---LYPLGGQTNLRIPQGTGQV----- 525
Db      510 SFDEAPFLKDLLEKIKDPSPEVLSALKALQHNGMLMVEDTVSSLSILHRIKRSADWC 569
Qy      526 -----MLDD-----RAYLVMEYSYSMTLF-----TCEIEMLAHV 556
Db      570 PVLKEAVRVLDDPRIIEGNPDLKAYI-----SMELLPVMTGRAPECELOMTSAI 621
Qy      557 VSTADVIQHCORVKEPID--LVHKVISTDLSIADCLLPITSRIYMLQRLTTVSPVVD 613
Db      622 TETLISQH-----PLTQMAKVLKAVLAKTSESD-LLGVAN-----EMLITTLV----- 664
Qy      614 VIASCNCLTVLAARNPAKWMTDL-----RHTGFAPFVHVSLSQWISAEQNA 664
Db      665 -IKNDAN-WDAATKANTLENVCDILSRQSSVDRDAAVVPSALQLOSMTESQHLHT 722
Qy      665 G-----GYGNLL-----MNSBOPQGEYGV 683
Db      723 AQSVYKLEPILLQAVTIOPEGVSDQPADBCLPVCVALGEFQKISGSLAEQEBGL--L 780
Qy      684 TIATLRL-ITTL-----VKGO-----LGSQSGQ-LVPC 710
Db      781 LLSLRRLFTTLTKCPDSTFKGEPWNNPEKMETTCCYRLLCRLDVAVISGASQGPLAPC 840
Qy      711 VMPFLKEMLPKV-----HKMRYNSHGVRQEGCLLELIHA----- 746
Db      841 FRSLMQPLQVHLNPPVLFKFLSLSWGNSN-LGQUDCRVSAILOTOALYVGAFLSS 899
Qy      747 ---ILNCHETDLHSSHTPSLOFLCISLAYTEAGQTVINIM----- 785
Db      900 QPVKTLNLLASDS--SPVVPFL-LVCVCS-GVCEVRBAIAYLQCLSGLVSPHYPIVEK 955
Qy      786 ---GIGVDTIDWMA-----AOPRSDGAEQOGQOLIKTVKLAFSVTNNVIRLK 832
Db      956 LLKSEBIIADSSYLTOALSKEFEAVSRKDKKLAAYEQLLQCLQSPFC----- 1006
Qy      833 PPSNVVSPLEQALSHGAGNNLLIAYLAKYTHKDBLPRLAIOQLKRLATVAPMSVYA 892
Db      1007 -PSYTSKTLRLQD--VHGEVLSVL-----PAVERLLBOCAPDSCTFLP----- 1050
Qy      893 CLGNDAAIRDAFLTRLOSKEIDEMIKWILEFLVAVETQGLIETFLNLEVKDSDGS 952
Db      1051 ---DEALL---LQLLSKSEMSAPLIVD-----PRCLIEVPIR----- 1083
Qy      953 KEFSGWMSCLHVALELIDSOQOQRYWCPRLIHRAIAF-----LHALMDRRDSAMV 1005
Db      1084 -----ALHTSARPYPIPSFOITALEQ----- 1105
Qy      1006 VLKTRP-----KFWENTLTPLF-----GLTSPSEBSEPILECLIMKICIEIY 1053
Db      1106 -ITPFPFTAIGDEKIQOKILISILPDLVGNKSPACQASINSVPKTIAYDCLVANLELP 1163
Qy      1054 V---VKGSLDQSLKOTLKPKFSEIKRFAVWGVVXSIAHVAVATEGSSCTSLILEYOMLV 1109

```

```

Db      1164 ADKQVATATVQOTRRSMRK-----TQDTSGA-----VPEBSVVS 1198
Qy      1110 AMRMLIIATTHADIMELTDSVVRQULFDVLDGTAKLLIWPASVNCILRSGMKCTLLI 1169
Db      1199 WPRVTLIL-----ELLQHKKKLRAQYLVRALPILRLRCLEPRAABQENIETVQILLIC 1253
Qy      1170 LLRWKRELSVDEILQPLTEILEGLVQADQOLMEKTKAKVYSAFVYLQMKKAKVDIP 1229
Db      1254 LNAVCKR---LSDEGPIIS--KDVLEBDKFME-----LVVQCVRVSEMP 1293
Qy      1230 Q-----YSQVLYANVCEILOEBVIA 1248
Db      1294 QTHRRALLLGLAGLPEPEKYLHNPITFTFMGANIMELDDTYSFOVIN--KTVOAVIPA 1351
Qy      1249 LPDQTRSLALGASATBDKDSMETDSCRSRRRDQDVCVGLHAKELCEVDEBDQSWL 1308
Db      1352 LI--KAHE--GSSQSGBNMET-----VVAQIINVFDALPHV 1385
Qy      1309 QVTRRLPILPILTTLEVS-----LRMKONLHFEATLHLTLARFOQGANAVAGA 1360
Db      1386 PEHRRPLISQWMTLPSRFLVIMLLFPQ--HTQ-----TSAGATGAKE 1432
Qy      1361 GITQS-----TC-----LPLSVQLSNGTAQTP-----SASRK--- 1390
Db      1433 AVERDDQFWILVCCBEVEKQULSLIKILOYIMTLPODBREABEKKKPRGRSAVKDET 1492
Qy      1391 -----SLDASWPGVVR--LSMSLMEQLKTLRYNPLPALPFGVGHQERTLOCL--- 1438
Db      1493 VSDILFVSVEITSGDLNFKRISISFMAQLAS--DGFVGVAQCEDI--TESTILOALQOD 1549
Qy      1439 ---NAVTVQSLA-CLEE-AD-----HTVGFILQLSNFMKWHFHP 1475
Db      1550 LLEVLRATYQAVACVEDNADKPTAKFWBALLSKSYDTLDVYNALLPMDTFI----- 1601
Qy      1476 QLMEDIQVNLGYLOACTSLSHSRKLOHTYQNRNGDLPBAVAVQVRPPSASAPSS 1535
Db      1602 TVMRGLVGN-----QLASVRKAMELLNNK-----LQOR----- 1630
Qy      1536 SKOPADTEASEQOALHTVQYGLIKILSKTLAAL---RHFTPVQCILLDQSLDLAEYN 1591
Db      1631 ---TKMLKEQIT-----ALHELIGTILSYGRSHROVTADE---ZEBELAINQOT 1673
Qy      1592 PLFALSTPTTFPDESVAPSGTLLATVNALNMLGEIDKKEPYLQAVGSTQAEGRFTL 1651
Db      1674 ALYGLKLCRNFQGDHKEEF--VPVANKAVELVADKDEKGVWGS-A-LICVAEVTSTL 1728
Qy      1652 KSLM-----FTMENCFTYL--ISQAMR-----YLDPVAVH 1680
Db      1729 KALAIPOHLRMPAVLDTLKERKOLLNBEIYLLSAVVALQASETLPHFISPYLDTLILQ 1788
Qy      1681 PRDQRMQKQELS--ELSTLSLSRYPRGAPSPATGVLPSPQKSTSLSKASPEQOE 1738
Db      1789 VTRLTLKRLRLTSCPOLSVRLASLSTL--ATKLPRVVLPIITKYCQSNVDAQORNL 1845
Qy      1739 PLIQLOVAFVRHMR 1753
Db      1846 PLMNILKEHISHQK 1860

RESULT 31
Q6MYH8
ID      06MYH8      PRELIMINARY;      PRT; 1818 AA.
AC      05-JUL-2004 (TREMBLrel. 27, Created)
DT      05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DR      05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE      Hypothetical protein.
GN      ORFNames-AFA683.010;
OS      Aspergillus fumigatus (Sartorya fumigata).
OC      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC      Eurotiota; Trichocomaceae; mitosporic trichocomaceae; Aspergillius.
OX      NCBI_Taxid=5085;
RN      (1)

```

RP SEQUENCE FROM N.A.  
 RX PubMed=14998527;  
 RA Pain A., Woodward J., Quail M.A., Anderson M.J., Clark R., Collins M.,  
 RA Foster N., Fraser A., Harris D., Larke N., Murphy L., Humphray S.,  
 RA O'Neill S., Petrea M., Price C., Radhikowitsch E., Rajadream M-A.,  
 RA Salberg S., Saunders D., Seegar K., Sharp S., Warren T.,  
 RA Denning D.W., Barrell B., Hall N.;  
 RA "Inserting into the genome of *Aspergillus fumigatus*: analysis of a 922  
 kb region encompassing the nitrate assimilation gene cluster.",  
 RI Fungal Genet. Biol. 41:443-453(2004).  
 DR EMBL: EX649606; CAF32025.1; .  
 KM Hypothetical protein.  
 SQ SEQUENCE 1818 AA; 202590 MW; 64FE189BFCB2974 CRC64;  
 Query Match 2.4%; Score 217.5; DB 2; Length 1818;  
 Best Local Similarity 17.2%; Pred. No. 86-05;  
 Matches 315; Conservative 301; Mismatches 699; Indels 515; Gaps 74;  
 58 SPSSAKVYKANKOVAS-----PKEL---GLRISKFLGLDE----- 90  
 67 SPSSAKVYKANKOVAS-----PKEL---GLRISKFLGLDE----- 126  
 91 EGVVOLLQCYLOEDRGR-----DSKTVLQDEROSQAL-----ILKI 129  
 127 RPAALILAFTEESTSLQSAAGIDMLRASLAGPALSEVLRQKAGDSDPLESENRLRL 186  
 130 ADYVVEERTCILRCVLLHLLTY-----FODERHYREYADCVDKLEKELVSKYRQFEL 184  
 167 RNLVLSSESHLKTARKLALSLHNSQNDNAPRAVHSGRIQSLKCAAFQOE----- 241  
 185 YKTEAPTW-----ETHGNLM-----TERQVSRWFVQCLREQSMLLEI 221  
 242 -KSESESRMSFVEACIKAIKORLSALEGGGWLGVASESEAVENMRITLIEVAVHQM 300  
 222 IFL-YTAAPENAPSDLV--LTRKEQSGFSGRQTNRLVDETMDFVDRIQYFALLIV 278  
 301 LFLQASAEIPTAELVSWLRVMDYSLFESVQVSHPCNPLRLPL-LQAFVALTTL 359  
 279 EGMNDI-----ESLHKCALDRRELHOFADGLICODMDCLMTPG-DIPIHAPVLAMNL 332  
 360 AFMKPLAMPSTIINKSTPSPGSPYFLSKGEIGC-LNEIFVTALESSTANPAAPSMGL 418  
 333 LRLHTLNP-----EETSSVVRKI-----GGTAIQLVFOYVLTLLQ----- 367  
 419 LHLTRRELALSDKEIRELEGFHSAVDSPQSNTPHNSGQASLSYBELLECARPTCA 478  
 368 --SLASGNDCTTSFACMCVGLSLFVLTSLEHLTLANOQ----- 405  
 479 EDSIALTLTSDAMKOTMFEITVALATNVGSTSANDILTERWTRVLLDLIRVRIYVEYS 538  
 406 -DIIDTACEVLADP-----SLPELFWGTEPT---SGLGIIIDSVCGM---FPHLSPL 451  
 539 PEIVGSVALILEGPTKLLCSQSLSVATDPKIMTNDPFLMDELPHILAKSRFFETVTF 598  
 452 LQLRALVS-----GKSTAKVYSFLDKMSFYNEL--YKRPDIVISHDG----- 495  
 599 LKLCALISKDMVYNEGIPOILSEWENMDFTQIYPLFEGYETIREDENNALVTLQPL 658  
 496 TLMRQRTKLYL-----PLGGQTNLRIPQGVQGVMDADRILVYRMSYSSTLTFCE 549  
 659 AIFPSSSTARQIMHDHPSNALVVTSSHPSSITLQVISEAKPAVIMWHEYSCLSYLSY 718  
 550 IEMTLHVSTADVIOHCORVKKPIIDLVHKVISTDLSINDCLLP----- 592  
 719 LEEWNEENGSSGVVEE-DNVAEIIIGLM-----ADLLVAWVQPTNEASGAKRILEVAS 772  
 593 -----ITSRIYMLQRLTTVISP-----PVDVIAVCVCLTVLAARPAKWTDLR 638  
 773 DGLARKSDIISLVVPIFERILHNIIGPRAGDSKVSVALACRFIALTMVLPGRW----- 828  
 639 HTGFLPFAHVVSLSQMIISABGNAGGYGNILMAN-SFQPOGEYGTATFALITTLV-- 695  
 829 -----FLIARSSILGSDG--KGGVMTAIVSATETVSGDYFPLLECVTLFQDVVD 876

QY 696 -----KQQLGTSQSGVPCVM-----FV--LKEMLPSTYHKRNSHG 731  
 DB 877 AASRAMVKPCNNLSGKRTAASEWTSVPTTHMRGILLNFAVMTVEVFNISNDIMRNAE 936  
 QY 732 VAEQIGCLILIEHAIILMCHETD-----LHSGHT-----PSLQPLCISLATEAGQVIN 783  
 DB 937 QLEINTTLAKTPEMIIYTTGTNATNATLDSKVTGVSPSSATYLLDVLARPGRDLPRP 996  
 QY 784 IMGIVDTITD-----NVMAAQPRSDAGEGGGQGLIKYKALFASVTNNVIRLKP 833  
 DB 997 ILRLIVDGLQTPATLHLRYLVLMKQVRS-----TLELTRILAEARYLEQ 1042  
 QY 834 PENVYSPLEQALSQGHAGNNLILAKYIYHKHDPALPRLALIOILKTL-----ATV 885  
 DB 1043 PSL--LEDOIFKASPLVTKVLSLDAY-----RLPVILLFILLISGALDSAN 1089  
 QY 886 APMSVYACIGNDAAIRDAFLTRLOSKIEDMRKIMLEPLTVAVEYTOGGLIEFL----- 941  
 DB 1090 EPPSLVGHLAGASSCLFLDVLSDQPKPLSDQKQLAVHNFSLTYSKQOMLAVYILGOS 1149  
 QY 942 ---NLKVGQSDGSK-----EFLGMSCLH-----AVLELIDSOQODRYWCP 982  
 DB 1150 SARQTMKQTDYEGGPTWKGTPFLQIALGMLNIBQMDLRAALSLIEFVSCAQENMPMATP 1209  
 QY 983 LLHRAALFHALMQDRDSAMLVLRTEKREMNLTSLPGLTSPSESTSESLTET-CA 1041  
 DB 1210 -----ELRKTPOFSSLVN--YVKLTKIPLVPDQIPATIA 1245  
 QY 1042 LIMKILCLEIYVYVKGSLDQSLAKDTLKKEFKEPAYSGYVKSIAVAVAEFGSSCTSL 1101  
 DB 1246 AVADLCAVYLHSAKMDQDRFTKTLIP-----LVFM--YSKD-AVDVSNVNSLHNL 1296  
 QY 1102 -LEYQMLVANRMLIIATTHADLMH--LDSVVRQOLFVLVDGTA----- 1146  
 DB 1297 KRNEMRYSGCKL-----ADFRRTSLQPRSGRGDYDYIQLGDKLSDYDFWAGTKN 1348  
 QY 1147 -----LLVPAVYVNCRLSGMKCTLLILLRQKRELGASVDILGPLETLE 1193  
 DB 1349 QGRABEPRRANINISLEAOS-----LHSMKFP-----ATEHCT 1384  
 QY 1194 GVLQADQOLMEKTYAKYFSAFTVLOKMKMKVSDIPQ-----YSQVLNVCE 1240  
 DB 1385 DFM-ADRE-VQKSNASVVGSL-----EANKIGVPEAIRIQRVREFAQLL--Q 1433  
 QY 1241 TLQF-----EVALPQOTRHSIALSGATEDKSMETDCCSRSHRDQDGVLYGLHIA 1294  
 DB 1434 RLVEAGAKGAEVFGILREVWMDALPTRRATYE-EALVNDDAEYR-----SVLVNVL 1484  
 QY 1295 KECVEVDEGDSMTQVTRRLPIPLTLTTLEVSIRMKONLHTEATLHLTLARTQGA 1354  
 DB 1485 ALQCHDSNPTTDTTSRIAEIPSDILVIEIKYVAVHAFBSLITTLHQDPKCEPD 1544  
 QY 1355 TAVAGITOSICPLLS-----VYQLSNGTGAQTPSASRKSADPSMPGVYRLSMSLM 1408  
 DB 1545 FALITALIQTSLQKNDRLYEHIIVHIBDNNTA-----RHANSLPSWADQAVSGPV 1598  
 QY 1409 B---QLKTLRYNLPALDPVGVHQRRTLOCLNAVTVGSL----- 1447  
 DB 1599 YGSLSTFLVYGLSTLPLALHIAV--EAVLTRISTCGLTNILQPRGPRDAPRLTYI 1656  
 QY 1448 -----ACLEADHTGVFILOLSNFMKEWHFHLPOLMRDIQVNLG-----YLOACTSL 1495  
 DB 1657 WTAGFLPLCLNLVHLYLTAEVAAALNQBEGOLKRALESFVADRSQSPSRRLCLSMASB 1716  
 QY 1496 LHSRKMLOHYLONKNGDELPSA--VAQRVQ 1523  
 DB 1717 AYSLALIS-PLINRFREAGPSAGVDAQSIQ 1745

RESULT 32  
 CAF32025  
 ID CAF32025 PRELIMINARY; PRT; 1818 AA.

AC CAE32025; (Tremblrel. 27, Last sequence update)  
 DT 13-MAY-2004 (Tremblrel. 27, Last sequence update)  
 DT 13-MAY-2004 (Tremblrel. 27, Last sequence update)  
 DT 13-MAY-2004 (Tremblrel. 27, Last sequence update)  
 DE Hypothetical protein.  
 GN AF683.010.  
 OS Aspergillus fumigatus (Sartorya fumigata).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocommata; mitospore Trichocommata; Aspergillus.  
 CC NCI\_TaxID=5085;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Pain A., Woodward J., Quail M.A., Anderson M.J., Clark R., Collins M.,  
 RA Fooker N., Fraser A., Harris D., Lark N., Murphy L., Humphrey S.,  
 RA O'Neill S., Petter M., Price C., Rabbittowitch B., Rajadream M.A.,  
 RA Salzberg S., Saunders D., Seeger K., Sharp S., Warren T.,  
 RA Denning D.W., Barrett B., Hall N.,  
 RT "Insight into the genome of Aspergillus fumigatus: analysis of a 922  
 kb region encompassing the nitrate assimilation gene cluster."  
 RL Fungal Genet. Biol. 41:443-453(2004).  
 DR EMBL, BX649606; CAF32025.1; --  
 KW Hypothetical protein.  
 SQ SEQUENCE 1818 AA, 202590 MM, 64FE189BFCB2974 CRC64;  
 Query Match 2.4%; Score 217.5; DB 2; Length 1818;  
 Best Local Similarity 17.2%; Pred. No. 86-05; Indels 515; Gaps 74;  
 Matches 315; Conservative 301; Mismatches 699; Indels 515; Gaps 74;  
 QY 58 SPSSAKRKANKDVSS-----PLKEL---GLRISKFLGDE----- 90  
 DB 67 SPSSAKRKANKDVSS-----PLKEL---GLRISKFLGDE----- 126  
 QY 91 EOSVQQLQCYOEDYRGR---DSVKTVDLDEROSQAL-----LKI 129  
 DB 127 RPAARLLAFTEEEESTLSQAGINDLRASLAGPALSEVLRQKADDDPSLEENRRL 186  
 QY 130 ADYVEERTCLRCVHLITV---FODERHRYVEVADCVDKLEKELVSKYRQFEL 184  
 DB 187 RNLVSEERSHLKTKARKLLASLHNSQNDNAPPAHVSGRIQSLRKGAAYFOE--- 241  
 QY 185 YKTEAPTV-----ETHGNLM-----TERQVSWFVQCLREQSMLEI 221  
 DB 242 -KSSESEKRSFVEACIKAIKORLSALBEGGMLGVAESEAENMRTYLIIEVLVQCM 300  
 QY 222 IFL-YVAVFEMAPSOVLV--LTKMFKEQGFSGRQTNRLVDETMDFYRIGYSALLIV 278  
 DB 301 LFLQAGASAEIFTAELVSWLWMDYFLESIVQVSHPCQNPRLILP-LQAFVALTVL 359  
 QY 279 EGMNDI-----BSLHKCALDDREHLQFADGILICQDMCLMTFG-DIPHAFLVLANL 332  
 DB 360 AFMKLPLAMPSTINKSTPTSPSOSFYFLSKGEIGQ-LNEIFVTAGLESKTANPAAPSMGL 418  
 QY 333 LRHTLNP-----EETSSVVRKI-----GGTAIQLANFOYVTLRLQ----- 367  
 DB 419 LHLTRRELALSKETRELEQFHSAYDSPOSNTPHNSGQASLSLYBELLECARPTCTA 478  
 QY 368 --SLASGNDCTTSTACMCVGLSFVLTSLBLATLGNQ----- 405  
 DB 479 EDSIALTLSDAMKOTMFETIVALATNGSTSAVDILTERMTRVVLLDLIRVRIYEVYS 538  
 QY 406 -DIIDTAEBVLADP-----SLPELFWGTEPT---SGLGIIIDVSCGM---FPHLSL 451  
 DB 539 PEIVSSVLAILEGPTKLCLSSQSLSVAIDPFWIMFNDPFLMDELPHILARSFPYEVVF 598  
 QY 452 LQILBALVS---GKSTAKVVSFLDKMSFYNEL--YKHKPHDVISHEBG----- 495  
 DB 599 LKLCALSLSKDMVNEEGIPOLISMEHNDFTFOIYPLERQGYETREDENALVLTQPL 658  
 QY 496 TLMRRQETKLY-----PLGGQTNLRIPQGTVGOMLDDRAYLVWREYSYSWTLFTCE 549  
 DB 659 AIFBSSTARQIMDHPNALVVTSSSHVPSSTLGGVISEAKPAVIMWYHEVYCLSLGXY 718  
 QY 550 IEMLLHVSTADVIGHQCRVKEPIIDLVKAVISTDLSIADCLP----- 592

DB 719 LBNMENGFSFGVEE-DNVAETIGLM-----ADLLVAVMVFPTNBSAGKRLIENVAS 772  
 QY 593 -----ITRIYMLLORLTTVISP-----PDVNASCVNCLTVLAARPAKWTDLR 638  
 DB 773 DGLARKSDIISLVDPFERIILHNIGPRAGLDSKVESVJACRFIALTMVLPGRVW----- 828  
 QY 639 HTGFLPFAHVSSLSQWISAEQWAGGVNLMN-SQPOGEVGTATPRLITLV-- 695  
 DB 829 -----PULARSSLSGSG--KGGVMTALVSAIVTSGDIPFLLECYKLPDQVDD 876  
 QY 696 -----KQGLSTOSQGLVPCVM-----FY--LKEMLPSYHKMYNSHG 731  
 DB 877 AASRAMVRCNNLSGKNTAASWTSVPTHMIGILNFPRTWVEVNSINDRFPNPE 936  
 QY 732 VRBOIGCLILHAIYMLCHETD---LHSGHT---PSIQFLICGLATAGQTYIN 783  
 DB 937 ORLEINTIATKFERMIYTYGTNDATGLDSKVGVFSSSATYLLDVLRPOSADLPFNP 996  
 QY 784 IMGIGVDIID-----VWMAAPRSDGABEGGQGLIKYKLAFSVTNNVIRLXP 833  
 DB 997 ILRLIVDGLQTPATLHLRYLVLMKQVRS-----TLELTRLEAARYLEQ 1042  
 QY 834 PSNVVSPLEQALSGHAGNNLIYVAKYIYHKDPALPRLAIQLKRL-----ATV 885  
 DB 1043 PTLT---LEDDLFASPLVLYKLSLDAY-----RLPVILFEILISGALDSAN 1089  
 QY 886 APMSVYACLGNDAAIRDAFLTRLOSIEDRIKMIYLFVAVETOPGLIELFL--- 941  
 DB 1090 EPPSLVGHLAGSSCLFVDSVGFQKPLSDQGLQAVHFPSTIYSKQOMLAAYILNGS 1149  
 QY 942 -----NLBYKQSDGSK-----BSLGWMSCLH-----AYVELDSQGDQVYMC 982  
 DB 1150 SARQTMKQTDYEGGTMGTPFLQIALGMLSNIGMDIRALASLSLEFVSCQENPMKTP 1209  
 QY 983 LLHRAIAFLHALMODRDSAMLVLTCKPKWENLTSPFLPTLSPSESTSEPSILNET-CA 1041  
 DB 1210 -----ELKTPQPFSSLVN--YVKLTKIPSLPVQDQIFATNIA 1245  
 QY 1042 LIMKIICLEYVYVKGSLDGLKDTLKFSIEKRPAYSGYKSLAVHAETESSCTSL 1101  
 DB 1246 AVVADLCAYVLSAADMQDRTFYKTLIP-----LVFW--YSKD-AVVDVSAVNTSLHANTL 1296  
 QY 1102 -LEYOMLVASARMILLIITTHADIMH--LTSVVRQFLVDVDOTKA----- 1146  
 DB 1297 KRFEMRYSGGKL-----ADFRRTSLQPSRSLGRDYVDIQLGDKLSYDFAWSGTKN 1348  
 QY 1147 -----LLLVPSVYNCRLTSGMKCTLLILLRQWRELSYDEILGPLETILE 1193  
 DB 1349 OGFAEFERANINLSLVAQVS-----LHSMKFF-----ATEICT 1384  
 QY 1194 GVLQADQOMKERTAKYVSAFITYLQMKMKVSDIPQ-----YSQVLAVNCE 1240  
 DB 1385 DFM-ADRE-VQKSMASVQSCF-----EANKVGPQAIPEIRIQGTVEFQALL--Q 1433  
 QY 1241 TLQF-----EYIALFDOTRISLALGATBEDKSNETHDDCSRSRRDQDQVCTGLHA 1294  
 DB 1434 RLVEAGAKGAVFGILRVVMDALRTTRATYR--EALVNDAAEYR-----SVANVFL 1484  
 QY 1295 KEICEVDEGDSWLOVTRRLPIPLFTLTLTSLVSLMKONLHFTETLHLLTLARTQGA 1354  
 DB 1485 ALQCHLDSNPRTPTSSRIARISBDLVLEIVTVVAHGRSLITYLHQDPKCEKXD 1544  
 QY 1355 TAVNAGITQSICTPLS-----VYQSLSTGTAQTPSARKSLDAPSPGVYRLSGLM 1408  
 DB 1545 FALTLAIQTSLOIKNADRLYEHIYTHIEDNTA-----BHAMSLFEMADQLAVSGDPV 1598  
 QY 1409 E---OLTLTKYVNPFLPALDVGVGHQERTVLCNAVKTQSL----- 1447  
 DB 1599 YGELSILFLVGLSTPLMAEHLAV--EAVLTRLSCTRLNTIIOQRGFGPDVAVRLYTI 1656  
 QY 1448 -----ACLEBADTVGFIIQLSINFMKMHPLPOLMRDIOVNLG-----YLQACTSL 1495

Db 1657 WTAGFLPLCLNLTLYVLTAEVAFLNQEGOLKRAESFVADRSGSPSRICLSMASE 1716  
 Qy 1496 LHSKRLQHYLQNKNGDGLPSA--VAQRYQ 1523  
 Db 1717 AYSLALIS-FLINRFREAGPSAGVADSIQ 1745

## RESULT 33

Q9T014 PRELIMINARY; PRT; 561 AA.  
 AC Q9T014  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
 DE Hypothetical protein AT4g38760.  
 GN Name=AT4g38760;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=1702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,  
 RA Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL035656; CAB38610.1;  
 DR EMBL; AL161594; CAB80539.1;  
 DR PIR: T06075; T06075.  
 DR Hypothetical protein.  
 SQ SEQUENCE 561 AA; 62921 MW; DBCCS12226FFCD1C CRC64;

Query Match 2.4%; Score 215.5; DB 2; Length 561;  
 Best Local Similarity 22.2%; Pred. No. 1.7e-05;

Matches 132; Conservative 99; Mismatches 244; Indels 119; Gaps 24;

Qy 86 LGLDEBQSVQIQYQEDYKGRTRDSVKTLQDE-----ROSOQLTL 127  
 Db 9 NWLDEIQSYILVERSMGEY-GTDSVAQELTQEFIDMSFLHNCQFVWYSECAFLL 67  
 Qy 126 KIADYYERTCILRCVHLHLYFQDERPRVVEVADCVKLEKLVSKYRQGFELYKT 187  
 Db 68 TLISLSPYS--NVLTEDTASVLTHTAPVTKVYKAYSHHTCI-----ICSVYAPRRESSIKE 119  
 Qy 188 EAPWETHGNLTEROVSRWFVQCUREOSMLTEIFLYAYFEMAPSDLVLTWKFKKGG 247  
 Db 120 EAVKILISQ--LER-----RQSSVLEDLSSCT-----PKMKGIS----- 153  
 Qy 248 FGSRTNRLVDETMDFVDRIGYSALIVEGMDIESLHKALDRRELHQFADGQ-- 305  
 Db 154 -GSYNSFKLAIVSAQSHSACRVOQLMILITELDMEN-----LLQWMDVGPVF 201  
 Qy 306 ---IC-----QDMDCLM--LTFGDIPRHAPVLAWL--LKTTLAPERTSSVVRKIG 350  
 Db 202 RSGGCVSVIVQEDATISSLNTISEVDAEGLVLAWVFLCLSSLGKESRPLMDID 261  
 Qy 351 GTAIQLNVQ-----YLTRLQSLASGANDCTTSTACQVYGLSFVLTSLRLTL--G 402  
 Db 262 HVSYVHQFBAASISYFLIEILQSNLNDPDEIGSHREVVTFISAFASFEINDQLEDE 321  
 Qy 403 NQODIIPACVLT-ADPSLPELFMGTE-----PTSGLAGIIDSVCAMPHLLSPLDQLR 456  
 Db 322 TLEILILILSKYQGEESLCCQFMDRKSFVVDGIFCLFLFLES---EPPFSAFIRLLS 378  
 Qy 457 ALVSGSKTAKVYEFLLDKMSFYNNELYKKRPHDVISH--DGLMRQRQPKLLYLGQGTN 514

Db 379 SLESGSWPACVYNFNDKSVGVSTLP-----DITSDSPADDSQVLETSRPLH-IGLSG 432  
 Qy 515 LRIPQGVQVM--LDDRAYLVREXYSYSWTLFTCEIMLTHVSTADVICHQCVKPI 572  
 Db 433 LVIPSWTRGRIARVISENTVLRMEYLSGLIIVLIRLNKYLIGNRRAP-----VT 485  
 Qy 573 IDLVKRVISTDLSADCLLPITSRIYMLQRLTVVISP---VDVIACVNCIT 623  
 Db 486 LELIRRMVTFENKAVCPISLNIHSFFYQGSYVNGKMSDVRVVDIICNSVRSIT 539

## RESULT 34

N188 YEAST  
 ID N188 YEAST STANDARD; PRT; 1655 AA.  
 AC P52593;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Nucleoporin NUP188 (Nuclear pore protein NUP188).  
 GN Name=NUP188; OrderedLocNames=YML103C;  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 [1]  
 RP SEQUENCE FROM N.A., AND INTERACTION WITH N1C96.  
 RA Zabel U., Doye V., Tekotte H., Wepf R., Grandi P., Hurt E.C.;  
 RL "N1c96p is required for nuclear pore formation and functionally  
 RL interacts with a novel nucleoporin, Nup188p."  
 RL J. Cell Biol. 133:1141-1152 (1996).  
 [2]  
 RP SEQUENCE FROM N.A., AND INTERACTIONS WITH N1C96 AND POM152.  
 RA Nehraas U., Rout M.P., Maguire S., Blobel G., Wozniak R.W.;  
 RL Pubmed=6682855;  
 RL "The yeast nucleoporin Nup188p interacts genetically and physically  
 RL with the core structures of the nuclear pore complex."  
 RL J. Cell Biol. 133:1153-1162 (1996).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX MEDLINE=97313268; Pubmed=9169872;  
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
 RA Jagels K., Iye G., Moule S., Odell C., Pearson D., Rajadream M.A.,  
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrett B.G.;  
 RL "The nucleotide sequence of Saccharomyces cerevisiae chromosome  
 RL XIII."  
 RL Nature 387:90-93 (1997).  
 [4]  
 RP SEQUENCE OF 1187-1205 AND 1611-1629.  
 RX MEDLINE=96095775; Pubmed=8522578;  
 RA Alchison J.D., Rout M.P., Martelli M., Blobel G., Wozniak R.W.;  
 RL "Two novel related yeast nucleoporins Nup170p and Nup157p:  
 RL complementation with the vertebrate homologue Nup152p and functional  
 RL interactions with the yeast nuclear pore-membrane protein Pom152p."  
 RL J. Cell Biol. 131:1133-1148 (1995).  
 [5]  
 RP CHARACTERIZATION, AND NPC SUBUNIT LOCATION.  
 RX Pubmed=10684247;  
 RA Rout M.P., Alchison J.D., Suprapto A., Hjertaas K., Zhao Y.,  
 RA Chait B.T.;  
 RL "The yeast nuclear pore complex: composition, architecture, and  
 RL transport mechanism."  
 RL J. Cell Biol. 148:635-651 (2000).  
 [6]  
 RP REVIEW.  
 RX Pubmed=12791264;  
 RA Suntharalingam M., Wente S.R.;  
 RL "Peering through the pore: nuclear pore complex structure, assembly,  
 RL and function."  
 RL Dev. Cell 4:775-789 (2003).





```

QY 1216 T-----VLQKEMKVSIDIPOXSOLVANCETLQSEVIALPDQTHSLALGSATEDKQSMET 1271
DB 1188 TDSASLQVYVWEIETSKAMGALITTVK-----NSTVL----- 1222
QY 1272 DDCSRSHRDQDQVCLGLHLAKELCEVEDGDSMLQV--TRRLPILPILTLLEVSILRM 1330
DB 1223 -----NDGPDVLVHEHFLKNTIDFGSDKQMFQIYLERIELEFYLSPKLSGKL 1271
QY 1331 KQNLHTEATLHLTLTLARQOGATVAGAGITQSICTPLLS-----VYQUSGTACTPS 1386
DB 1272 LKEEKTEIEMNKKTFTLFKSGEIDFIKXIGSLKKNFRPLRSVLVLEKIVSSGDRFTEL 1331
QY 1387 ASRKSU---DAPSPGVYRLSMISLMEOQLTKLRNPLPRLADFGVQOERTLCUNAVRT 1443
DB 1332 ISDQLEFPFELVSKVYLILSLQINK-----CSTRGSTDHTTQIVN--- 1377
QY 1444 VQSLACLEADHTVGTILQSNF-----MKEMHFLPOLMRDI----- 1481
DB 1378 -----LR--DNTQDILLILSLFKITVNVPNSKNFVTLASSLNBVGTLKVIILNYSAN 1429
QY 1482 -----QVNLGYLQACSTLSHSRKLQHYLQNKRGDGLPSANVAVQVRPSPSAS 1530
DB 1430 LRRINDEPLLGQTTTLFISLCSIEPIAAKL-----NSGLYSVLE----- 1471
QY 1531 AAPSSSKOPAADTEASEQOQLHTV--QYGLKILSKTLAALRHP---TPVCOIL--LDQ 1583
DB 1472 -SPLSVAIQGGDIKPFSPRLNHNINSGLISV---LILLSQGINVLPEPTCLFVYRFGK 1527
QY 1584 SUDLAENLFLALSFTPTPDSVAPSGTLL---LATVVALNMG--EL-----D 1629
DB 1528 QIKSTLYN-----GDNDLAVSSSLIKETNQLVLLQKMLNLTNOELFIQPKNSD 1577
QY 1630 KKKPEPLTQAVGSTQAGRTTKSLMTFMENCFYLLISQMRXYLNDPAVHR----- 1682
DB 1578 DQGEAVELVIGLDSBNDKRLSALSKF-----LTHPKYLSRII 1617
QY 1683 ---DKQMKQELSELSTLLSLSR 1704
DB 1618 PTLLEQQLDEBSRLR-FVKGISR 1642

RESULT 35
OCCT5 PRELIMINARY; PRT; 1548 AA.
AC Q6C1T5;
DT 01-OCT-2004 (Tremblrel. 28, Created)
DT 01-OCT-2004 (Tremblrel. 28, Last sequence update)
DT 01-OCT-2004 (Tremblrel. 28, Last annotation update)
GN Chromosome P of strain CLIB99 of Yarrowia lipolytica.
ORFNames=YALIOF13519;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CX Saccharomycetiales; Dipodascaceae; Yarrowia.
NCBI_TextID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RG GENOLEYURES;
RA Lafontaine I., de Montigny J., Marc C., Nevegilise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Batney S., Blanchard S., Beckerich J.M., Beyne E., Bleykaesen C.,
RA Boissiere A., Boyer J., Caticolico L., Confantolero F., de Daruvar A.,
RA Despons L., Fabre B., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantrege F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerreb A., Koszul R., Lemaire M., Leaur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolaki M., Oztas S., Olier-Kalogeropoulos O.,
RA Pellerau S., Pottier S., Richard G.F., Straub M.L., Suleau A.,
RA Swenne D., Tekela F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenon-Meyer M., Zivanovic I., Bojotin-Pukhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpell C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;

```

```

RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: CB382132; CAG78184.1;
SQ SEQUENCE 1548 AA; 172980 MW; D1198B6F5FB89E3 CRC64;

Query Match 2.4%; Score 212; DB 2; Length 1548;
Best local similarity 19.4%; Pred. No. 0.00014;
Matches 306; Conservative 240; Mismatches 532; Indels 496; Gaps 78;

213 REOSMLBETIYLYAYFEKAPBDLVLTMMFEGGFGSKQTRNHLVDEMDPVDRIQYF 272
DB 233 RQVEILKAVNGYFFDENPESADV-----KQMFESID-NLPFGESBAFC-----Y 280
QY 273 SALIVEGMDIESLKCA--LDDR--RELHQPADGLICODMGLMTRPGDIPHH----- 323
DB 281 SILISLVLDLQMFQDVEDSDLDKVDVDPETDEPIPDNVYFKDPATLTAVANNIIQEMNER 340
QY 324 --APVLLAWA--LIRHTINPEETS-----VVRKIGTVALQNLV 358
DB 341 LMSPIVLAMSHVLAQVTFROSETNDRLVLPFIQVSADONGSPYIVCTQLANGSLNRGV 400
QY 359 PQLYRLRLQSLA--SGANDCTSTACMCYGL--LSFVLTSLHLTLGANOQDIIDTACVYL 415
DB 401 LAVTQSAVAGLAKSAEHTMTVMISLCRASFRVLPVLTBAELVAVYLGPPYQV---CRYL 456
QY 416 ADPSLPETFWGTEPFGIGIILDSV---CGMPPHLSPLQLRLALVSGSKTPAKKYSPFL 472
DB 457 LENTLK-----IRDMALIALSRFPPIBEPFLYIMLAM--DERTAFTEYQMR 501
QY 473 DKM-----SFNELYKHKPHDVISHDGLTNRQTKLLYPGCGTNLRIPQSTGVQVWL 527
DB 502 TLMQQLPGFOXYTESSTDTNIVLTIGV-----PVFP-----BFIYPTGISTLVV 548
QY 528 DDR-AVLYRWESYSWTLFCIEIMLHVS--TADVIQIQCRKPIIDLVHKYIS--TD 583
DB 549 GDNIPPLVMWATFERFGAVY---LGLKLSQLSGRFAL-----IKMKRKYISVD 595
QY 584 LSADCLPITSRYMLQLRTTVVSP--VDVIAS-----CYNCLTVLAARPAKV 634
DB 596 AEVADDELRSAS---MMEEDIVYIAKLVYVAYANQHLFVATESKFTALCRIDPDIW 652
QY 635 TDLRHTGFLPVVAHPVSSLSQMSIAEGNAGCYGLNLSNPOQEGYVTTAFLRLITTL 694
DB 653 PFLSRAGIL-----AONGRPFGAAMVIGSTEIIVNGEYSLTALDLADTL 697
QY 698 VYKQUGST-----OSQGLVPCWMEVVLKEMLPSTYHKMYNKHQVREOIGCLLELIH 745
DB 698 VENNVAASIDSKVSAKEQSDVLAHCLTRYI--DVRESCGYWRYVNDQ-----VELMS 749
QY 746 AILMLCHETDHSHTSISQFLCISLAYTERAGQVINIMIGVDTIDMWAAPRSDGA 805
DB 750 KSVAMFOGT-----VHYAANMNVVPLRA---SAEYIILGFLTSIGR 789
QY 806 EGQGGQGL-----IKTVGLAFSVTNVIRLK-----PPSNVVSPLBQALSGHAGN 853
DB 790 DLGQRPRLTCLDSTSVEMVNASISLEPLTKIVEARTNLELAPVLEKQFLASPQ----- 843
QY 854 NLIVLVAKYIYHKGDPALPRLAIOILKRIATVAMSVYACIGNAAIRDAFLRLQSKI 913
DB 844 -LVELYOQFPYHK-----KVIDLMLALVDSWVDCQSPLLHLGQFHSOL----- 888
QY 914 EDNRKIMLEFLTVAVETQGLTEFLANLEVKDQSGSKFSGMSCLAVLELDSQ 973
DB 889 -----LKSLSAMEN-----NLE-----TETVIGICAFPAVVD--GSS 922
QY 974 QQDRYKPCPLHRAAIALFALNMQRDSAMLVRTPKKMENTISPLF----- 1022
DB 923 RQDQ-----LSILVTRKDSILVLETRALSKQVSSVFPVAVIKAMALAR 968

```



```

QY 1023 GTLSPSTSEPSILSTCALIMKICLEIYVYKSLDQSL-----KDTLKKEFKERFAY 1078
D 969 NTWSVTYKTSKORLTK-----LVGRWV--EVDSDVPQLPEKIRSPKSSKRYTVAARQOR 1023
QY 1079 MSGVYSKLAHVH-----AETEGSSCTSLLEYQWLVAMFLLIATTHADIMHLDTS 1130
D 1024 FSA---AAAVQYIABELPRDPTVGA-----LISKW-----GSKMLKLTSS 1062
QY 1131 VVRRLQFLVDLGTALLVPSVNCRLGSKCTLLTLILRQ-----WKELGSV 1181
D 1063 FSSIQLYRSLHG-----NLIRFPAKMMLELFQFKQDPEYGSFVVDLEDM 1111
QY 1182 DEIGCP-LTEILEGVLAQDQOLMEKTKAVPSAFITVQMKEMKSDIPQYSQV----- 1235
D 1112 NAVFEPSELGEPESEFAENANLINS-----LVNSSTELS--LAWQQLLVAVID 1155
QY 1236 LNVCTLOEVALPDQTHSLALG-----SATEDKDSMETTD 1273
D 1156 VKIABSLGDIVVAME-----ISLGMIAFVEGVYNNALFVIFHILKSPBLKVDVTE- 1209
QY 1274 CSRSHRDQDQVGVGLHLAKELCE-VDEGDSMLQVTRRL--PILFTLLTLEVSLEK 1330
D 1210 -----GLKLCWALLESSTVDSFLSSLSGTSRLKYPVLRG-----ISILV 1247
QY 1331 KONLHPTFA---THLL--LTILARTQCATVAGAGITQSTICPLISVYQUSTGTACT 1384
D 1248 RKALNPEDMAQVHGISDIYLAJ-----GSYALANA-----IQYDVSHKG----- 1289
QY 1385 PSASRKSIDAPSPWGVYRLSMLEBQLKTLRYNPLPEALDVGVAHQERTLOCLNAVTV 1444
D 1290 --HSEITVDVDFLIEDLRIMVLLRNLL--SFENTCVPVSLAOPAL---KIDDTGSIRSL 1342
QY 1445 QSL-ACLEBAHTVGFIIQLSNFMKEMHFLPOLMRDIOVNLGYLCOACTSLHSRKMLO 1503
D 1343 LSLVYSQGEIDPVFG-----ELTLLFILE-----LLQNDQJSE 1375
QY 1504 HYLONKNGDGLPSAVAQVQRPSPASAPSSSKOPADTEASBQOALH--TVQY---G 1557
D 1376 QVVLN-----GLVGAALTE-----APISRE-----IQKSLHOGRLYYIWMRG 1412
QY 1558 LKLIKSLTAL-RIFTPVQIILDQSLDLAEYNLFALSTFTPTPDEVAPEGTILA 1616
D 1413 ILPIMLTVLVRGRIISPEITML-----LSF-----FOQLISTALRNLM 1452
QY 1617 TVNVALNMLGELDK 1630
D 1453 PVTISLSFVDELDQ 1466

```

## RESULT 36

```

O6CNP1 PRELIMINARY; PRT; 1606 AA.
AC O6CNP1.
DT 01-OCT-2004 (TREMBLrel. 28, Created)
DT 01-OCT-2004 (TREMBLrel. 28, Last sequence update)
DT 01-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Strain NRRL Y-1140 chromosome B of strain NRRL Y-1140 of Kluyveromyces
lactis.
GN ORRNames=KLLAOE11033g;
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RG GENOLEVURS;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marcq C., Neugebille C., Talla B.,
RA Goffard N., François L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykassen C.,
RA Boletane A., Boyer U., Catolico L., Contantolero F., de Daruvar A.,

```

```

RA Despons L., Fabre E., Fairhead C., Ferry-Dimazet H., Groppi A.,
RA Hantreay F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Kozul R., Lemarié M., Lesur I., Ma L., Müller H.,
RA Nicoud J.M., Nikolski M., Ozias S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potter S., Richard G.F., Straud M.U., Suleau A.,
RA Swenne D., Tekela F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zentou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,
RA Winker P., Souciet J.L.,
RT "Genome evolution in yeasts.",
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RA Genoscepe;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL, CR382125; CAG99535.1;
SQ SEQUENCE 1606 AA; 182208 MW; 0C5837B000AF2363 CRC64;

Query Match 2.4%; Score 212; DB 2; Length 1606;
Best Local Similarity 17.4%; Pred. No. 0.00015;
Matches 339; Conservative 295; Mismatches 606; Indels 712; Gaps 90;

QY 76 KELGRIKSGFLGDBEQSVOLQCTYQEDYGRDSDVXTVLQDERQSAALIKADYYE 135
D 38 EELVHLSRFTSVENTDSVG-----KQLQNGY--SYVATVEDELANBELSLANVHQE 90
QY 136 -----ERTCLRCVLLHLYTFODERHPVREYVADCYDKKEKELYSKR 178
D 91 CLRVISQTRTFPPDQPMKQDITVQKILH-----ER-----NCV--LDITVLNLR 133
QY 179 QQFEL-----YKTEAPTWETHGNLMTREV-----SRMFOQLRQSMLEIIFL- 224
D 134 DFPVLKSMQOLVYKNRS--TWISAILLVKNCVALISABONDSYKSLMSYDILYIN 192
QY 225 ---YAAFEH-APSDLVLTTFMFKQSGRQTNHLYDETMDPFVDRIGYFSALILVEG 280
D 193 LNLNLAIPSLNQPVLEIITKYE-----LLEFQ-----YEA--ILKDN 230
QY 281 MDISLHKCALDRREHLQPAODGLICQDMCLMT-----FGD----- 319
D 221 KDVP-----EIVMANSLMI--MDIIMTGYNTSBDITDVKSSFPNDATPR 277
QY 320 -----IPHHAPVLA--WALL-----RHTLPBETSSVVRKIGTA 353
D 278 LLDLIINTDPPALIIYVYNSFIIYSKSYLLEDPENNLFPVQVFNQSPISKIKSPALTA 337
QY 354 IQLNFOYLTRLQSLASGNDCTSTACMCYVGLSFVLTSLEHTLGNQDIIIDTAC- 412
D 338 ENENVFERISKVSQIF-----KQAFPAALILSFITPSLYFVPISIKTSMITVLSNAPT 393
QY 413 ---EVLADPSLPELFMGTEPTSGIILDSVCGMPHLSBLLQLRLALVSGSKTAKV 468
D 394 HYVEKFLTNKEPEKFF-----LIKTKAPLKESLIPMINITS- 432
QY 469 YSFLDKMSFVNELYKXP-----HVISHEBDGTLMRQTPKLY 507
D 433 ---NAEFANFEWELPTTYQEVVLSLEVDLADAHSTDLISIKSETPA- 480
QY 508 PL--GGQTNLRIFQSTGVQVM-----LDDRVLVREYYSWTLF-----TCEIEMLH 555
D 481 FLBPSGNVLPRIKQTKAKILPSVIND----VMLELYKYNKMSLGRVLYQNC- 532
QY 556 VVSTADV-IQHC---QRYKPIIDLHVKVISTDLSI---ADCLPITSRIYMLQRLTYI 608
D 533 YLSTSSAGVQDSVDQEMILSMDLISSVIGSGISVEKADILLQNST--YLIEDYIFDI 590
QY 609 SP-----PVDYIACVNCVLTTLAARNPAKVTDLRHGFLEFVHAPVSSLOMISA 659
D 591 FKYFQOSILRADYVIDICMLNLSSIDDFQFVSHL-----ARSELDR 636
QY 660 BGMNAGYGNLIMNSQPOGEGYVTAIFRLITTYV-----KQQLSGTQSG 705

```

```

Db 637 YG-KSGLISSVLTGMEIPLNGNYSITIKIKLAQQLVLESLTETEPENIKKELISKPFQ 695
Qy 706 GLVPCWFWLKEMLPSYHKWRYNSHGVREJGCLILBILHAILMCHETDLHSHTSLQ 765
Db 696 HFI-----HVESYQYNNYRSIAEXELGTMLTTLFTTVNNYIGID--PSNPETK 745
Qy 766 FLCICSLAYEAGQVINIMIGIVDTIDMWAAPRSDAGAGQOG-----QLIK 816
Db 746 ITKVLSSAANDITNRFLSSDSPOVRTYSLIHVLSSNSLSLEALSIGNEVPINKNOCILK 805
Qy 817 TVKLASVTNNVIRLK-----PNSNVSPLEQALSQHGANGNNLIIVLAKYIKHHPAL 871
Db 806 SYELA-----NLLISVRAMLQLPSTLESQI-----XGN-----L 835
Qy 872 PRLAQLKRLATAPNSVYACLGNDAAIRDAFLTLQSLIEMRIKIMILETVAVE 931
Db 836 PKL-----IDIVA-LHND-----LRNPI-----TKMTHLVRAWAE 866
Qy 932 TOPGLIELPLNLEVKDSDGSKFSLGMSCLHAYLEIDSQODRYWCPFLHRAIAF 991
Db 867 NPSLSLSYL-----GQKSKMLFNSILYDLASTIDYQ-----PLKSLYRF 907
Qy 992 LHALMODRDSAMLVRT-----XPKFMENTSPFLFGLSPSTSEPSILLETGALI 1043
Db 908 FSSVMEGKODGFSLIVTGERTISVSSPKKSE-----PSNGASSTSL--SL 953
Qy 1044 MK-IICELIYVVVKGSLDQSLKOTLKFKFSEIKRPAYMSG-----YKSLAVHAYET 1093
Db 954 KKNVLKLDCT-----PESVASRLLE-AISYAFNSVAAARNNTADKEFIDFLVKLEKF 1005
Qy 1094 EGSSCTSLLEYQW-LVSAMRMLLIATTHADIMHLDTSVVRQLDGLTALILVPA 1152
Db 1006 EDLAPDSVEATIKLSGHRVLRQARIARIFA--LHLFSSSDSTIMNANGPQ---LVDI 1060
Qy 1153 SVNCLAGSKCTLLILRQMKRELSVD-----EIL--- 1185
Db 1061 VKRIFSIDGYDKTLHLLAEFKKIMPHLDIEKFAISPLRIGKAYQANIYDIELDQF 1120
Qy 1186 -----GPILEIEGVLOADQOLMEKTKAKVFSAFITV---QMKMKXSDIP 1229
Db 1121 CCEKRGKNGSTGFREQVKSINLOYSFQISAASGALLTAPTKTPKLOASVYD 1180
Qy 1230 QVSQVL-LNVCELTQEBVIALFDQTRHSLALGSATEDKOSMETDSCRSRHRDQRGVCV 1288
Db 1181 IVTQLALNVHEGVAPIFS-----EIVL 1204
Qy 1289 LGHLAKECEVDDEGSMLOVTRRLP--LPLTLTTEVSLMKQ-----NL 1334
Db 1205 LRLELVFYLY-----SFLQTSMKIPEDLETLTLOLALIKSKEVDYGINAQGNKN 1258
Qy 1335 HFTL-ATLHLLTLARTOGATAVAVAGITGICLPLSVQSLSTNGTAQPPSARSKL 1392
Db 1259 QYRPILRSILMLITLTKPY-----PMSEERS 1285
Qy 1393 DAPSWGVRYSLSMT-MEQLKTLRYNPLPEALDFVGHQERTLOCLANVRVQSLACT- 1450
Db 1286 D-SLEVEVELAFSSKGVNIIISILFSF-----IASNKGKVYIPNLADKIQDILL 1337
Qy 1451 -----EEDHVTGVFLQLSNFKEMW-----FHLPOLMRD-----IOVNIQYLCQ 1490
Db 1338 SLPTKIKELNPSSTSFNAVLASSINBEGTLKSLNLTYYSSYLFERSHDEPLFAELTSLFSIE 1397
Qy 1491 ACTSLHSHKMLOHLYONKNGOGLPSAVAVQRYORPPSASASAPSSKOPADOTHEASBOQA 1550
Db 1398 LVT-----IEEVAQKIINDGLFSVFLE-----SPISVSIQSGAPPELSPR-- 1438
Qy 1551 LHTVOYG-----LKLISK-----TLAALRHPTPD 1575
Db 1439 LHNINSNGIISIILOLSKFGSGLSEVCLFVSYFESKOJDLAVGSSSLSLITQAYIYS 1498
Qy 1576 VQO-IILDDSLDLAEYNPLFALSFTTPTDSSEVAVSFGTLLATVNVALMAGELDKKCP 1634
Db 1499 TSOIILLOKMLDLKLEYORLSTVN-----GSKLKVID-----EERI 1533

```

```

Qy 1635 LTOANGLSIOAG-----TRTLKSLMFTMENCPLYLLISQAMRYLRDAVHPR----- 1682
Db 1534 IELPGLDSSSEKNDLANVTKHL-----THPKYNSRIIAT 1570
Qy 1683 --DKORWKOB-----LSSELTLSSSL 1702
Db 1571 TPDKQOMEDDEVNRQNLHSHITQISLQSSSL 1602

RESULT 37
ID 09LUS3 PRELIMINARY, PRT, 1837 AA.
AC 09LUS3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Similarity to unknown protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroside II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asanizu B., Kotani H.,
RA Tabata S.
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RL clones."
RL DNA Res. 7:31-63 (2000).
DR EMBL:AB023044; BAA97383.1;
SQ SEQUENCE 1837 AA; 206399 MW; 45B6B47B0E7095 CRC64;

Query Match 2.3%; Score 205; DB 2; Length 1837;
Best local similarity 18.7%; Pred. No. 0.00056;
Matches 377; Conservative 273; Mismatches 708; Indels 662; Gaps 93;

Qy 48 LEGSLTYKPPSPSSAKRYKA-----NKDVASPLKEGLRISKFLGDEQS 93
Db 40 LONLSPFPKPSDQAQVQSKERLPDLPLSLDQDIA-----ISLKSDELHNLNIDS 94
Qy 94 VOLLOCVIOE-----DYRTPSVTVLODERGSQLIKIDVYYS 135
Db 95 VRLVSSNOEGMGRDPLEIORLATGMYTGRRLTSLY----- 135
Qy 136 BRTCLRCVLTMLTYFODERHPYVEVADCVDKLEKELVSKYRQOFBEILYK---TEAPTW 192
Db 136 ---TLRAVV-----LDEGLRBDLI--ADIGLLELLEAGLRQRLITLIKELANREDPTG 185
Qy 193 -----ETHGNLTEROVSRWFVQCLREOSMLLEIFLYAYAYEMADSDLLVLYTK 241
Db 186 LGGPLCEERYLIDSRGALVERBAV---VQ--RERLIGHCVLSTLVDRPGSKVDVDIY 239
Qy 242 MEKEGFGSRQNNRLVNTMDPFYDRIGY---PSALLVSGMDLSLHKCLDDBREIH 298
Db 240 ILKD-----NAAQLTEGNDTSSQITFSLFLITTPVSDALS-----RLSDKSSSN- 285
Qy 299 QPAQDGLICQMDCLMTFQDIP---HHAPVLTAMALLRHTLNBPETSSVVRKIGGTAI 354
Db 286 -ISQDASFRTPQDVIWASGSDPTADGRTGIRLAMAN--HLM----- 325
Qy 355 QLVNFQYVTRLLQSLASGNDCTSTA-----CMCVYGLLS-----FVLTSLEHNTIGN 403
Db 326 -----LIHGIQGMOTITSTASTDWMGHICGCSLEISFKNVFGFLDNV--LRTAY 374
Qy 404 Q-QDIIIPACVLT---ADP-----SL-----PELPKGTBPTSGIGI 435
Db 375 QVKSRSKDNAMSVLNSYRTSDPLDGSMTGERSDRPLPFLISLMEFKPELILSGND----- 427
Qy 436 ILDSVCGWPPHL-----LAPLQLALAVSGSKTAKKVVGFLLDKWSP-- 477

```

428 -----VLMTFVNAGEDHTNFKTLVAFLEMLCTLASTOGASKVYELRGTSFRSIGMPTL 483  
QY 478 -----YNELYKHKPHVDIshedgTLMRQTPKLIYPLGGOTNLARIPOGTGQVNLDBRA 531  
Db 484 FDCIRIDEKFKOS-----LQTAGAMMP-----BFLGDAK 514  
QY 532 YLVWEMYSASWTLFTCEIMLHVSTADVIQHOCRVKPIIDLVHKYISTDLADIADCL 591  
Db 515 ALVAVY-----LNVLOKVENGNERPFR-KAMFPDIEPPFKLIGIE----- 552  
QY 592 PITSRIYMLQRLTTVISPPV-----DVASCVNCLTVLAARNPAKWTDLRHTGFLPF 645  
Db 553 -----NIPYLKALRKRTIAAFVNVPEMR-----DSIWAFLQYDLPRV 592  
QY 646 VAHVPYSSISQMIASGNMAGGYGNLNMSEOPQGYGTIAFLRIITLVKQGLSTOSQ 705  
Db 593 VGSQVQKSDOSSOVYDMQFE-----LNEVEAREQYPSITISPLNIMNLAGE-KDVNDR 646  
QY 706 G-----LVPCVWFVLEKMLPSYHKMRVNGHVBEQICLIELIHALNL-----CHETDLH- 757  
Db 647 GRRAYSDBC-----EKY-----QLVAVACLOHFMHILSMYDIOESDLQGF 685  
QY 758 -----SSHPTSPLOF-LCICISLA--YTEAGQTVINIMGI---GVDTIDWVMAAQPRSD 803  
Db 686 TEHPHFLVSLTSSLOQOLPIIILKDFMSSKALYRNLMILOVGNISIERLSKT--- 742  
QY 804 GAEQOGQOLIKYVKAFAVSTNNYIRLK-PPSNVVSFLBQALSOGHAGNNLIAYLAKY 862  
Db 743 -----YKILEKAVQSLIELILLVFEKOLVSDWRBLYOPLDIILLODHOIILALLEY 796  
QY 863 IYHGDPALEPLATOL-----KRLATVAPMSVYVACLDNDA-AIRDAFLTRLOKIEDM 916  
Db 797 VRVDSLPQIOSSIKIMNIRCSRLVGLVPMILKI---DAANSILBYAACLSERLEBG 852  
QY 917 RI-----KXMLEPLTVAV-ETOPGLIBLFNLLEVYDSDGSKESPLGMSCTIAVVL 967  
Db 853 EVVENSOCODLVLMQLLVNDINIRAPSIITHLILKFDLPDAPRGVGLPKPHYSGLKXIL 912  
QY 968 ELISQOQDRYWCPR---LHRAIAFLHALWQDRDS--AMLVIRTPK--FWENLTP 1020  
Db 913 EMLEKLPY-----PDINFLFEFGFOLCEMLADPLTGPTWDLISKKYOFLOHLDLT- 966  
QY 1021 LFGTLPSP-----SESEPSILETCALIMKITCLBY-----Y 1053  
Db 967 -TGAATLPKRSQSQALRISLSHQRAWMLKLALIAHTSGSSSAHLEACQSLHLPRE 1025  
QY 1054 VVKGSLLDQSLKDTLKKFSIEKFAVWSGVKSLAVH-----AETEGSSCTSLIE 1103  
Db 1026 VTEANERFSSSTYQDGLD--YAGTSSISKSKALALEILOFRPDASMOLPOIVSSLK 1083  
QY 1104 YQMLVSAWRMLLIATTHADIMHLDTSVVRQLF-----LDVLDGTAKL----- 1148  
Db 1084 YDSLYB-----DILGNRDTSVSGSIYYERSGDRILDLSFSPNLMQKLSHG 1130  
QY 1149 --LYPASVNCRLPSMKCTLLIILIRKWK-----RELQ 1179  
Db 1131 PPLVDSFENVAVELSEVRRTIOQLKMGKYNRNLEBOAOLHMLAGSQIVEVASCRIS 1190  
QY 1180 SVDELIGLTLIELSEVLQADQOLMEKTYAKVSAFI-----TVLOKMEKMS----- 1226  
Db 1191 SLDN-----RSEILIRILDSASASAPDCSLMAFVLVQVALLTCIAKARDRFRSQALS 1246  
QY 1227 -----DIPQYQVLNVVC-----ETLOEVAL-----FDQTRHSLA 1258  
Db 1247 SDVTJCLDVWVWKHLSGTACHSVFLKVAALIRHSSSLRRROYALLLSYQYQCHMIA 1306  
QY 1259 LGSAT-----EDKDSMETD---DCSSSRH---RDQDGYC----- 1287  
Db 1307 LDVPTSVVQFLNLNODSEBDLDIOIKIDKQADLABANFFIIRKBAQGLIDVYKASOGS 1366  
QY 1288 -----VLGHILAKELCEVEDDQSMLOVTRRLPIFLTLITLLEVS-----LRMKONLH 1336  
Db 1367 EFGKTIISLYBALVCIHBERYFLSQDSKGRPIRSCLSISNISYQDGHLLBSQORACT 1426

QY 1337 TEATLHLLTLART--QOGATAVAGAGTOSICLPLSVOLSTNGTAQTPSASRKSUDA 1394  
Db 1427 LEAEIALALRISHKYGSGGVLPMSGALBHIA-----SCRAISIRKNNR-----RVDKL 1477  
QY 1395 PSMFG-----VYRLSMSIMEQLKTL-----RYNLPBALDPVGHQERTLOQ 1437  
Db 1478 QSDVGNVQKORTIITVNLRVFAL-TSLVTSSEFPBERRNKIVADVAFIKHQGLPQL 1536  
QY 1438 LNAVTVOSLACLEADHTVGFILQLSNFMKEMHFLPQLMRDIOVNLGYCOACTSLIH 1497  
Db 1537 LREDFTQADMLIMEQIILAVGILSKWMPFEENDGQFVQGLFDMMSKLI-FASPISKILS 1595  
QY 1498 SR-----KMLQ-----HYLOKNG-----DGLPSVAORVORPPSAASPSK 1537  
Db 1586 QKSSBLKLSQRFSLTYLTVTKNSLRLOVSDSLDST--KLROPTLLILSLSHV 1653  
QY 1538 QPADTSEAEQO-ALHTVQ-----YGLKILSKTLAALRHFP----- 1574  
Db 1654 TDSLERAEEKSLHMKIRINELSRQVDALIKICD---SQEVYPSDNIHKRYIAM 1709  
QY 1575 -DVQI-----LIDQSLDLAEVNP-LFALSFTPTPDSVAPSEGTILATVVALNMG 1626  
Db 1710 VEMCOIVGNBDQITLLQLAEHYLNIILHLODRSVSSNERGSGSKSHIQOEVTDLCG 1769  
QY 1627 ELDKKKEPLTOAVGSLTOABGRTLKSLMFTMNCPLL 1666  
Db 1770 KL-----SPTDRLALNEV-----SLIDRERSIIVL 1798

RESULT 38  
QY 07153 PRELIMINARY; PRT; 1278 AA.  
AC 071153  
DT 01-OCT-2003 (TREMELrel. 25, Created)  
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE SI:ZC146F4.2.2 (Novel protein similar to human BAR28) (Fragment).  
GN Name=SI:ZC146F4.2? (zebrafish) (Danio rerio).  
OS Brachydanio rerio (zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Garner P.;  
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL732629; CAB17602.1; -.  
FT NON TER 1  
SQ SEQUENCE 1278 AA; 143523 MW; E6C9FC81B77EB1A9 CRC64;  
Query Match 2.4%; Score 197.5; DB 2; Length 1278;  
Best Local Similarity 18.4%; Pred. No. 0.001;  
Matches 241; Conservative 180; Mismatches 401; Indels 487; Gaps 59;

QY 677 PQGEVGTIAFLRIITLVKQGLSTOSQGLVPCVMFVLEKMLPSYHK----- 724  
Db 6 PEKMETTCCYLRILCRFLFVAVIGASQGPAPCFRSLMOPDLOHNLNBPVLFKFLSL 65  
QY 725 WRVNSHGVRBQICGLIELHIA-----ILNLCHETDLSSHSTPSLOPLC 768  
Db 66 WGINSN-LGDDLDKRVSAIILQTOALYKKAFLSQPVKTLNLASDS--SPVPSL-LVC 121  
QY 769 ICSLAYTEAGQTVINIM-----GIGVDTIDWVMA-----A 798  
Db 122 VCS-GVCEVRBAALAVAVOCLSGLVSSPYHPLVLEKLSSEHIIADSSVLTQALSIFYEBA 180  
QY 799 QPRSDAGSGQGOGLIKTVKLAFAVSTNNYIRLKPPSNVVSFLBQALSOGHAGNNLIAY 858  
Db 181 VSRDKNKKLASVQLOCLQSPFC-----PSYTSKTLRALOD--VHGSPVLSV 228  
QY 859 LAKYIYHGDPALEPLAIAQLKRLATVAPMSVYVACLDNDAAIRDAFLTRLOKIEDMI 918

```

Db 229 LL-----PAYERLEQCAPDSCFLP-----DEALL-----LOLLSKFSMSA 267
Qy 919 KWMILEFLTVAETOPGLIELFLNLEVKDSDGSKFSLGMSCLHVALELIDSOQDRY 978
Db 268 PLLVKQ-----PRCLEVFIR----- 282
Qy 979 WCPPLHRAIAF-----LHALMODRDSAMLVLRKP-----KFMENLTSPLP- 1022
Db 283 ----ALHTSARPYPTIPSPFOITALEQ-----ITKFPFALIGDEKIOQKILSIFD 328
Qy 1023 ----GTLSPSETSEPSILETCALIMKICILEIYV-----YKGSIDDSKDTLKKSIFERK 1075
Db 329 LTVGNKSPACAGINSVFKTIADVCELVANELLIPADKORTATVVOQTRSKRK----- 382
Qy 1076 FAYMSGYVSLAVHVAETBSSCTSLLEYOMLVSAMRMLLIATTHADIMELTDSVVRQ 1135
Db 383 ----TODTSGA-----VPBESVSWPRVTLIL-----ELLQHKKCLKRAQ 418
Qy 1136 LFLDVLDTKALLLVPSAVNCLRGSKCTLLILKRMKRELSVDEILGPTLELGEV 1195
Db 419 YLVPLPFLNLSRCLERPAABQENIETKOLILCLANVCK-----LSPBGPIS---KDV 471
Qy 1196 LQADQOLMEKTRAKVPSAFITVLQMKEMKVSIDIPO----- 1230
Db 472 LEBDKFNM-----LVQCVRVSEMPQTHHALLLGLAGIFPEKVLHNM 518
Qy 1231 ----YSOLVAVNCEITLOBEVIALPDQTRHSIALGASATEDKDSMETDCC 1274
Db 519 PIFTFMGANIMRLDDTYSFOVIN--KTVOAVITPALI--KANE--GSSSGSEGHMET-- 568
Qy 1275 SRSRRDQDGVCLGLHLAKELCEYDEBDGSMLOVTRRLPILPTLTLEYS----- 1327
Db 569 ----VVAQIIVFDALPHVPHRRRPLISQMLSTIGSPRFLVLM 610
Qy 1328 -LRMKNHPTENTALLTLARTOQATAVAGAGITOS-----IC----- 1367
Db 611 LLLFKQ--HVTQ-----TSAGATGAKKBAVERDQDFWILVCCFEYKQJL 657
Qy 1368 LPLLSVYOLSTNGTACTP-----SASRK-----SLDAPSPGVTR--LSMGL 1407
Db 658 IKILDYIMTLPODRERAPKKRGRSAVKQDETVSDLIJSVETHSGKLRHFKFISISF 717
Qy 1408 MEOLKTLRYNPLREALDFVGNHOBERTLOCL-----NAVRTVQSLA--CLSE-AD-- 1454
Db 718 MAQVLAS--DSFVGKADCEDI--TESTLOALQODLVEVLRKIQAARCEVDNADKPTAK 774
Qy 1455 ----HTVGFILQDSNFKEMHFLPLQMLRDIQVNLGYLQACTSLLSHRK 1501
Db 775 FWRALLSKSYDTLDKVNALLPMDTFI-----TMRGLMGN-----QLASVRK 817
Qy 1502 LOHYQNKRGDGLBSAVAVQVRPSPASABSSSKQPAADTBASQOLHVVQGLKI 1561
Db 818 AMELLNKK-----LQOR-----TKMKQJTT--ALLET 843
Qy 1562 LSKTLAAL-----RHFTPDVCQILIDOSPLAEYNFLALSFTTPPDSVARSFGTLAT 1617
Db 844 IGTLLSIYGRSHROQTAGE-----EBELAINQTAIYSLKLCRNPGSDHKEF--VPV 895
Qy 1618 VNVALNMLGEIDKKEPLTOAVGLSTOAGRTLKSLM-----FTM 1659
Db 896 LNKAVELVADKDEBKVMGSA--LTCVAHVSTLKAALAIQHLRLMPLVDTLKERKOL 953
Qy 1660 ENCFILL--ISQAM-----YLRDPAVHRDKQKQKQELSS--ELSTLLSSISR 1704
Db 954 NNEIYLLSAVVALQASSETLPHRISPYLLDTILOVTRTLARRLTSCGSLSVRLASIS 1013
Qy 1705 YFRGAPSPATGVLSPQKSTSLSKSPESQEPILQVQAFVHMQR 1753
Db 1014 TL--ATKLPRAVLIPITTKYCSNVDKQONRLSPMLNKHSHMK 1059

```

RESULT 39

```

07T152
ID 07T152 PRELIMINARY, PRT, 1336 AA.
AC 07T152;
DT 01-OCT-2003 (Tremblrel, 25, Created)
DT 01-OCT-2003 (Tremblrel, 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel, 25, Last annotation update)
DE ST:ZC146F4.2.1 (Novel protein similar to human BAP28) (Fragment).
GN Name=ST:ZC146F4.2;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Actinopterygii; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Cyprinidae; Danio.
OC NCBI_Taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Garner P.;
RL Submitted (JUL-2003) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL732629; CAE17603.1; -.
FT NON TER 1
SQ SEQUENCE 1336 AA; 150326 MW; AA949557P21ACBCE CRC64;

Query Match 2.2%; Score 197.5; DB 2; Length 1336;
Best Local Similarity 18.4%; Pred. No. 0.0011;
Matches 241; Conservative 180; Mismatches 401; Indels 487; Gaps 59;

Qy 677 PGEYGVITAFRLITTLVKGQSGTOSGGLVPCWFLKEMLPYHK----- 724
Db 6 PERMETTCCYLRILCRLEFDVVISGASQGLAPCFRSLMQPLLOVHNEPVLPRFLSL 65
Qy 725 WRNSHGVRQIGCLILELHNA-----ILNCHETDLSHTPSLQPLC 768
Db 66 WQYN--LGDQDCVSAALQGLALYVGAFLSSQPVKTLNLAADS--SPVPSL-LVC 121
Qy 769 ICSLAYTAGQTVINIM-----GIGVDTIDMVNA-----A 798
Db 122 VCS--GVCEVRRALAVLQCLSGLVSPYHPLVEKLKSSBEILDSSVLTQLSKYBBA 180
Qy 799 QPRSGAGGQGGOLITVTKLASVTNNVIRLKPSPVNVSPLEQALSQKAGNNLAV 858
Db 181 VSRKQNKKLASVEQLQCLQSPFC-----PSYTSKTLRALD--VHGSPVLSV 228
Qy 859 LAKYTHKHDPLPLAIOQLKRLATVAPMSVYACIGNDAAIRDAFTRLOSRIEDRI 918
Db 229 LL-----PAYERLEQCAPDSCFLP-----DEALL-----LOLLSKFSMSA 267
Qy 919 KWMILEFLTVAETOPGLIELFLNLEVKDSDGSKFSLGMSCLHVALELIDSOQDRY 978
Db 268 PLLVKQ-----PRCLEVFIR----- 282
Qy 979 WCPPLHRAIAF-----LHALMODRDSAMLVLRKP-----KFMENLTSPLP- 1022
Db 283 ----ALHTSARPYPTIPSPFOITALEQ-----ITKFPFALIGDEKIOQKILSIFD 328
Qy 1023 ----GTLSPSETSEPSILETCALIMKICILEIYV-----YKGSIDDSKDTLKKSIFERK 1075
Db 329 LTVGNKSPACAGINSVFKTIADVCELVANELLIPADKORTATVVOQTRSKRK----- 382
Qy 1076 FAYMSGYVSLAVHVAETBSSCTSLLEYOMLVSAMRMLLIATTHADIMELTDSVVRQ 1135
Db 383 ----TODTSGA-----VPBESVSWPRVTLIL-----ELLQHKKCLKRAQ 418
Qy 1136 LFLDVLDTKALLLVPSAVNCLRGSKCTLLILKRMKRELSVDEILGPTLELGEV 1195
Db 419 YLVPLPFLNLSRCLERPAABQENIETKOLILCLANVCK-----LSPBGPIS---KDV 471
Qy 1196 LQADQOLMEKTRAKVPSAFITVLQMKEMKVSIDIPO----- 1230
Db 472 LEBDKFNM-----LVQCVRVSEMPQTHHALLLGLAGIFPEKVLHNM 518
Qy 1231 ----YSOLVAVNCEITLOBEVIALPDQTRHSIALGASATEDKDSMETDCC 1274
Db 519 PIFTFMGANIMRLDDTYSFOVIN--KTVOAVITPALI--KANE--GSSSGSEGHMET-- 568

```

QY 1275 SRSHNRDQRCVYGLHLAKELCEVDEBDGSMQVTRRLPILTLTLEVS----- 1337  
 DB 569 -----VVAQIIVHVPDALPHVPERHRLPILSOLMSTLSPSFLWTLM 610  
 QY 1328 -LRMKONHFEATLHLLTLARTQOGATAVAGITOS-----IC----- 1367  
 DB 611 LILFKQ--HTQ-----TSAGATGAKEAVVERDDDFMLVCCERFVEQJLTL 657  
 QY 1368 LPLISVYQSTNGTGAQTP-----SASRK-----SLDAPSPGVTR---LSMSL 1407  
 DB 658 IKILQYMLTLPDREBAPEKKKPRGRSAVKDETVSDLIFFVEVTHSGDLMHFKFISISF 717  
 QY 1408 MEOLIKLRNLEPEALDFVGNHETLQCC-----NAVTVOSIA-CLEB-AD----- 1454  
 DB 718 MAQULMS--DGFQGVADCEDI-TESTIQAQODLLVEVLYTQVAVACVEDNDKPTAK 774  
 QY 1455 -----HTVGFILQLSNFKEMHFLPOLMEDIOVNLGYLQCACTSLHSRKM 1501  
 DB 775 FMRALLSKSYDTLDKVALLPMDTFI-----TVRGMLGN-----QLASVRKK 817  
 QY 1502 LQHYLQKNGDGLPSAVAGVQVPPSAASAAPSSSKOPADTEASEQALHTVOYLTKI 1561  
 DB 818 AMELLNKK-----LQOR-----TKWLKQIT-----ALLEL 843  
 QY 1562 LSKTLAL-----RHPTDVOQIILDQSDLAENFPLASTFTTPDSEVAPSGTLLAT 1617  
 DB 844 IGTLLSTVGRSHRQVTAOE-----EESLAIHQALSLKLLCRNFGSDHKEBF---VPV 895  
 QY 1618 VVVALNMLGELDKKCEPLTQAVGLSTQAEGRTRTKSLIM-----PTM 1659  
 DB 896 LNKRAVELVADKDEKKNWGS-LLCVAEVLSTIKALAIPLQHLMPAVLDTLKERDOL 953  
 QY 1660 ENCFYLL--ISQNR-----YIRDAVPRDKQKQKQSLG--ELSTLSSLSR 1704  
 DB 954 NNEIYLLSAVLAQRASLTLPHPISPYLLDTLLQVTRLLTLARLITSCPLSVLASISS 1013  
 QY 1705 YFRGASSPATVLPSPQKSTSLSKSPSQEPLQVQAFRHMR 1753  
 DB 1014 TL--ATKLPVRVLPITTKCYCQMVDAQNRSLPLMLKHSIMDK 1059  
 RESULT 40  
 QYVMB8 PRELIMINARY; PRT; 2067 AA.  
 AC QYVMB8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE CG11943-PA.  
 GN ORFNames=CG11943;  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_taxid=7227;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adame M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoekins R.A., Gale R.F.,  
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,  
 RA Abell J.F., Agbayan A., An H.J., Andrews-Pfankuch C., Baldwin D.,  
 RA Ballew R.M., Baer A., Baxendale J., Bayraktaroglu L., Beasley B.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Broctier P.,  
 RA Buttle K.C., Butman D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pallos B., DeCher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fierla S., Fleischmann W.,

RA Foeller C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harrie M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howard T.J., Mel M.H., Ibegwan C.,  
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kechum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matcel B., McIntosh T.C., McLeod M.P., Mpherson D.,  
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weinstein J.,  
 RA Williams S.M., Woodruff, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RA "The genome sequence of Drosophila melanogaster."  
 RL The genome 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426065; PubMed=12537558;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise B., Hodgson A.,  
 RA George R.A., Hoekins R.A., Laverly T., Muzny D.M., Nelson C.R.,  
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Taber P.B., Tan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,  
 RA "Finishing a whole-genome shotgun: release 3 of the Drosophila  
 RT melanogaster euchromatic genome sequence."  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.B.,  
 RA Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Berman B.P.,  
 RA Beltenkova B.R., Celinker S.B., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.B.,  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review."  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426071; PubMed=12537574;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426072; PubMed=12537575;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426073; PubMed=12537576;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426074; PubMed=12537577;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426075; PubMed=12537578;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426076; PubMed=12537579;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [11]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426077; PubMed=12537580;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426078; PubMed=12537581;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426079; PubMed=12537582;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426080; PubMed=12537583;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [15]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426081; PubMed=12537584;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [16]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426082; PubMed=12537585;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [17]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426083; PubMed=12537586;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [18]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426084; PubMed=12537587;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [19]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426085; PubMed=12537588;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [20]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426086; PubMed=12537589;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [21]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426087; PubMed=12537590;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [22]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426088; PubMed=12537591;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [23]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426089; PubMed=12537592;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [24]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426090; PubMed=12537593;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [25]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426091; PubMed=12537594;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [26]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426092; PubMed=12537595;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [27]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426093; PubMed=12537596;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [28]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426094; PubMed=12537597;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [29]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426095; PubMed=12537598;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [30]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426096; PubMed=12537599;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [31]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426097; PubMed=12537600;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [32]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426098; PubMed=12537601;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [33]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426099; PubMed=12537602;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [34]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426100; PubMed=12537603;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [35]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426101; PubMed=12537604;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [36]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426102; PubMed=12537605;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [37]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426103; PubMed=12537606;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [38]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426104; PubMed=12537607;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [39]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426105; PubMed=12537608;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [40]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426106; PubMed=12537609;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [41]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426107; PubMed=12537610;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [42]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426108; PubMed=12537611;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [43]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426109; PubMed=12537612;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [44]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426110; PubMed=12537613;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [45]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426111; PubMed=12537614;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [46]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426112; PubMed=12537615;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [47]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426113; PubMed=12537616;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [48]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426114; PubMed=12537617;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [49]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426115; PubMed=12537618;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [50]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426116; PubMed=12537619;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [51]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426117; PubMed=12537620;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [52]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426118; PubMed=12537621;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [53]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426119; PubMed=12537622;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [54]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426120; PubMed=12537623;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.B., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [55]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426121; PubMed=12537624;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,

QY 30 LRELSEOTELANLKHMRRLLEGLSYKPPSSSAEKVANKDVASPLKEGLARISKFLGLD 89  
 DB 48 LRNPSPRESEKSRNLLRNALNMGV-----PMQHSKRTKMSQDLA-----DEAVILSDMFDD 98  
 QY 90 EEOVSQVLOOCTYLOEDTR---GTRDSVKTUQDEROSQALILKADYYEER---TCILRC 143  
 DB 99 EGFAYELL-CTAOROOKHNPGLSRGVAVL-----LYYDERKAIISCTLRD 142  
 QY 144 VHLIL-----TYFODERHPRVAYAD-CVD-----KLEBELV----- 174  
 DB 143 MPQVSGVSWNTLPRGKITGLVNTYASLVDGSGILRLQLLEBMDREKCAMLTTRRA 202  
 QY 175 ---SKTRQO---FEELYKTEAR---TWETHNIMTEROVSRWFVOCLEOS----- 216  
 DB 203 FSGKQIONQVLAGYEDIOQALMALFNWSAQRL-PRHAIILMHQILANKKHADGAMN 260  
 QY 217 ---MLLEIFLYAYAEEMASDLVLTAKMKEGFSGRQTNHLDVETMDPFVDIGYRS 273  
 DB 261 DVTLLIMLALVAY-----DTSMLLVTEBPEHTRLRPIFS 297  
 QY 274 ALILVEGMDIESLHKCALDRRELHOFAODGLICQMDCLM-LTFGDIPIHNAVLLAMAL 332  
 DB 298 DEBFAE-----CFLEB-----LYAOSMOAPRLNAIIVASFG-----LTLAS 334  
 QY 333 LRHT-LNPEETS-SVVRK---IGGTAIQANVFOYLRLLQSLASGNDCTSTACMCVYG 387  
 DB 335 LNHAPLOQATAIISIINRDEMILDEALGAQVFFHSLLE-----KOLVYST-QFFYR 387  
 QY 388 LLSFVLTSELHTLNGQODIID-----TACBVLADPS---LPELFMTERTPSGLGILL 437  
 DB 388 RVHLIIT-----DFIDPMNSKVBELRGRADESSHTIISFLEGELEBP----- 429  
 QY 438 DSVCGMFPHLS--PLLOLRALVSG-----KSTAKV--Y 469  
 DB 430 -----PULDSNFBELMLCVAKMGDPRVTIRLCNEVWPGDPNGSTAFKXTSSVSLF 482  
 QY 470 SFLDKMSFNELYKAKPHVISHEDGTLNRROTPKLLYVGGQTN-----LRIP 518  
 DB 483 KXISLAS-----DLRPO--TLPRK---AYLKMSIGLRTDFSGARCAPMMLRVP 524  
 QY 519 QGTGVQVMDLDRAYLVREMEYSYSWTL-----FTCEI-----EMLHVASTADYV- 563  
 DB 525 QWATG-----GKAVSMWHFT--TLGNYTSMKNDFTNIGMSGETIYTRSTPRAIT 576  
 QY 564 ---OHORVKEPIDLV-HKVISDLSIADCLPITSRIYMLQRLTTVISPPYDVILAS 617  
 DB 577 QREASHLAVAMGIIQAVAEHDEVSR-IMICDHPNMOTPOVLGLVACAT---PLFKAE 631  
 QY 618 CVCNCLTVLA-ANRPKV-WTDLRHTGFLPFVAHPVS-SLSQMTSABGMNAGGYMLLANS 674  
 DB 632 ILHTLAALAKSKETARVIMFHLASQIIPV--PVSRSYAQCSLLEBMS---QIEBRS 684  
 QY 675 EOPQGEVVTIAFLRLITLVKQGLSTOSOG-LVPCWPFVJKEMLPS---YHKMRNS 729  
 DB 685 EQ-----YMLTGILDLTLMTTNPKSISGAGQRPAYEGYIKSINSVLLKFNRAKV 740  
 QY 730 HGVBEOIGCLILELTHALINTCHETDLHSHTPS-LOPLCICSLAYTEAGOVINIMIG 788  
 DB 741 PSEKQVGAQCLKLLYLL-----ATYRPSAMDFL----- 770  
 QY 769 VDTID-----WMAAQPSDDAE-----GQGGQGLIKTYKL----- 820  
 DB 771 -ETVDEBPYPGFHVMQOLKSEMLQLLCEIVEVRRLDYNRQGGKLEBESLYALL 829  
 QY 821 ---AFSVTNVIRLKPPSN---VVSPLQAL-----SOHAGNNLIVLAKYIKH 867  
 DB 830 ILBALKONAFPEHSAANCPILLSGANMLDLNPSRKPDRVLTNI---KTV--TY 883  
 QY 868 DPALPLALQILKELATVP-----MSVYACLANDAAIIRDAFLTRLQSKI-----E 914  
 DB 884 NSMLPRHALAIKILASVTLQPNVSTQILSMYGQGSNEKLEIROGFVACLEMEVCGKHD 943  
 QY 915 DMRIKWML-----BFLTV-----AVETQPLIELF 940

DB 944 DDLLOQALANHVPLFGEDDLDNEREMSGEREYITISQELQAGDGKLESKPSILQ 1003  
 QY 941 L-----NLBYK-----DSDGSKFSIG-----MWSCLHAVLELI----- 970  
 DB 1004 LKEALIKFLFEMTLSQLPFPVYVFLFGDVULAFPMANGHQLAIEMQSSGVNSVVLLEX 1063  
 QY 971 -DSQODRKYCPPLHRAIA-----FLHLMQDRDSAMLV-----LRTK 1010  
 DB 1064 MKQHSNDKYC--HTARIYERLYHPLHGICANRRTEILRYFLTCSDPLNHLNLSL 1120  
 QY 1011 P-----KFWENITSPLEGT----- 1024  
 DB 1121 PRORHEDHVLAMGHILNCVSDVKLAAXGQMTFRQMCDDILILIGMERSSHGLME 1180  
 QY 1025 ---LSPPS-----ETSPSIL-ETCALIM 1044  
 DB 1181 LQHSILISOPSSFLAMDVLPAQGSISAVAHGASASLGAFTGVKLPISLQETSQGLH 1240  
 QY 1045 KIICLEIYVVKGSIDQSLKOTLKKFSEIKRPAVSGYKSLAVH---VAFEGSCTS 1100  
 DB 1241 VTRLDILVLRAGTUSO-----POLPEPDGHLITQLRDCSABAGANSHAN 1288  
 QY 1101 LLEYQMLVSANRMLLIATTHADIMHLDTSVY---RRQLFDVLDGFKALLVPAZYNC 1156  
 DB 1289 LINV-----RKYIYIIT--DELMVQSIISAGQRAI-----STREMLHNAVNI 1332  
 QY 1157 LRLSMKTTLLIILRQKRELGSVDEILGPLYEIL-----EGVLQADQOLMEKTKARVP 1211  
 DB 1333 NRVRTORCA-TIAPMAV-----GOLVQVLESNMDAVLPAFORROH----- 1373  
 QY 1212 SAFITVLOMKEMKYSDIPOYSQVLVANCETOESVILFPQTRSL--ALGSATED--K 1266  
 DB 1374 -IIDIVKULIKQPL-----QPIIEISVOYSFVLLILANLRCCYQAEQSPDLAAR 1427  
 QY 1267 DSMETDCCSRHRDQBDGVCVLGLMLAKELCEVDEBDS----- 1306  
 DB 1428 DLSNGNGNAINDSQVVALC-----LGRRAIGTSGDGGSGGRDGGSGNSMLRPIIK 1482  
 QY 1307 ---WLOVTR-----RLPIPLITLTVLBSLBRKONLHTEATLHLLITLAPQOGATA 1356  
 DB 1483 NLVOMINISEVKSQKLIINYSILNCLRIAKRLRTEHLE----- 1523  
 QY 1357 VAGAGITOSICLPLLSVYQ---LSTNGTAQTPSARSLDAPSWPVGVRLSMSIM----- 1408  
 DB 1524 -----YQETLISROESARTYKQGRDD-----RLRKMAAREV 1558  
 QY 1409 ---EOLKTLRYNLEBALDFGVGHQERTLOCLNAVTVOSIACLEBADHTVPILQLS 1464  
 DB 1559 GTFGBKLIIDYICH---DAVTGHVDCRMALACIDMISBLOAVSTLCDPVAASRGYLGHL 1614  
 QY 1465 NEMKEMFHL-----POLMRDIQV---NLGYLQACSTLSHSRRKLOHLYONKMGDL 1514  
 DB 1515 ESLDQSTALCGILOVPDHLRPLUYVESRMAFLTRANANSVGARLL-----TEOALGV 1669  
 QY 1570 LSNMKVYDQDPDKSSSEIKNNEPOTPLPSIDR-----FRSILPLPSLCLDALVNS 1720  
 DB 1572 FTFDPVQIILDOSLDLAEVNFPA-----LSFTTPPD 1604  
 DB 1721 LGP-----RNSAQAQVNLFLAHLIDMVEAMLRSATPPYMD 1755

Search completed: December 30, 2004, 15:13:50  
 Job time : 341.065 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: December 30, 2004, 15:00:06 ; Search time 55.809 Seconds

(Without alignments)  
2083.096 Million cell updates/sec

Title: US-10-719-385-2

Perfect score: 9007  
Sequence: 1 MIRKSKTSLVSPCRSSREL.....PESQEPALQVQAVRMQR 1753Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PCRTUS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	216	2.4	1131	4	US-09-248-796A-19110
2	147	1.6	2954	4	US-09-150-867-1
3	146	1.6	2549	4	US-08-265-967C-1
4	146	1.6	2549	4	US-08-305-790B-2
5	144.5	1.6	2549	4	US-08-471-112A-3
6	144.5	1.6	2549	4	US-09-950-634-3
7	143	1.6	2549	4	US-09-538-092-1112
8	143	1.6	2549	5	PCR-US95-06722-112
9	140	1.6	3878	4	US-09-914-259-11
10	136.5	1.5	919	4	US-09-270-767-44725
11	136	1.5	2802	3	US-09-542-331-1
12	136	1.5	2802	3	US-09-510-791-1
13	131.5	1.5	147	4	US-09-270-767-33392
14	131.5	1.5	147	4	US-09-270-767-48609
15	130.5	1.4	1383	4	US-09-540-236-3580
16	129.5	1.4	1826	4	US-09-198-452A-113
17	129	1.4	3210	4	US-09-538-092-1154
18	129	1.4	3248	4	US-08-353-700-1
19	129	1.4	2938	5	PCR-US94-00198-3
20	128.5	1.4	2938	5	PCR-US94-00198-3
21	127.5	1.4	3433	4	US-09-538-092-1136
22	127	1.4	2470	4	US-08-265-967C-2
23	127	1.4	2470	4	US-08-305-790B-3
24	126.5	1.4	1106	4	US-09-248-796A-15028
25	125.5	1.4	756	3	US-09-085-199B-9
26	125.5	1.4	1064	3	US-08-726-214-8
27	125	1.4	914	3	US-09-085-199B-4

28	124.5	1.4	1824	3	US-09-228-246-2	Sequence 2, Appli
29	123.5	1.4	1090	3	US-09-085-199B-5	Sequence 5, Appli
30	123.5	1.4	1824	3	US-08-680-327-3	Sequence 3, Appli
31	123.5	1.4	2818	3	US-09-542-331-2	Sequence 2, Appli
32	123.5	1.4	2818	3	US-09-510-791-2	Sequence 4, Appli
33	123	1.4	2749	4	US-09-385-222A-4	Sequence 78, Appli
34	122	1.4	903	3	US-09-228-986-78	Sequence 4, Appli
35	122	1.4	903	4	US-10-101-464A-78	Sequence 4, Appli
36	122	1.4	1958	4	US-10-028-946-4	Sequence 2, Appli
37	122	1.4	2054	4	US-10-028-946-2	Sequence 4, Appli
38	122	1.4	3066	3	US-08-952-127-12	Sequence 12, Appli
39	121	1.3	2487	2	US-08-328-254-6	Sequence 6, Appli
40	120.5	1.3	2987	2	US-08-970-269A-29	Sequence 29, Appli
41	120.5	1.3	2987	3	US-09-407-562-29	Sequence 29, Appli
42	120.5	1.3	3672	2	US-08-832-445-12	Sequence 12, Appli
43	120.5	1.3	3672	2	US-09-356-540-12	Sequence 12, Appli
44	120.5	1.3	3801	2	US-08-822-445-10	Sequence 10, Appli
45	120.5	1.3	3801	2	US-09-396-540-10	Sequence 10, Appli
46	120.5	1.3	3959	2	US-08-970-269A-30	Sequence 30, Appli
47	120.5	1.3	3959	3	US-09-407-562-30	Sequence 30, Appli
48	119.5	1.3	2930	3	US-09-417-822-2	Sequence 2, Appli
49	119	1.3	2713	5	PCR-US96-01735-1	Sequence 1, Appli
50	119	1.3	3119	1	US-08-246-982A-16	Sequence 16, Appli
51	119	1.3	3119	1	US-08-453-265-16	Sequence 16, Appli
52	117.5	1.3	2020	1	US-07-551-531-2	Sequence 2, Appli
53	117.5	1.3	2818	1	US-08-510-284-1	Sequence 1, Appli
54	117.5	1.3	2818	1	US-08-411-389-2	Sequence 2, Appli
55	117.5	1.3	2818	2	US-08-449-933-2	Sequence 2, Appli
56	117.5	1.3	2818	3	US-07-966-049A-2	Sequence 2, Appli
57	116.5	1.3	1251	4	US-09-252-991A-17263	Sequence 17263, A
58	116.5	1.3	2871	4	US-09-538-092-136	Sequence 938, App
59	116.5	1.3	3433	4	US-09-091-501B-10	Sequence 10, Appli
60	116	1.3	585	4	US-09-655-479A-20	Sequence 20, Appli
61	116	1.3	1190	1	US-08-337-600A-2	Sequence 2, Appli
62	116	1.3	1190	3	US-09-048-887-2	Sequence 2, Appli
63	116	1.3	2485	5	PCR-US94-00198-1	Sequence 1, Appli
64	116	1.3	2485	5	PCR-US94-00198-2	Sequence 2, Appli
65	115.5	1.3	1179	4	US-09-792-024-89	Sequence 89, Appli
66	115.5	1.3	1441	4	US-09-252-991A-28143	Sequence 28143, A
67	115.5	1.3	3056	4	US-09-360-416-2	Sequence 2, Appli
68	115	1.3	1110	4	US-09-270-767-42166	Sequence 42166, A
69	114.5	1.3	1077	4	US-09-412-210-1	Sequence 1, Appli
70	114.5	1.3	1077	4	US-10-121-911A-1	Sequence 1, Appli
71	114.5	1.3	2186	2	US-08-822-445-2	Sequence 2, Appli
72	114.5	1.3	2186	3	US-09-396-540-2	Sequence 2, Appli
73	114	1.3	1299	3	US-08-460-900C-62	Sequence 62, Appli
74	114	1.3	1299	3	US-08-674-509B-48	Sequence 48, Appli
75	114	1.3	1299	3	US-08-954-698-48	Sequence 48, Appli
76	114	1.3	1299	4	US-09-639-695-62	Sequence 62, Appli
77	114	1.3	1299	4	US-09-448-188-48	Sequence 48, Appli
78	114	1.3	1299	4	US-08-954-188-48	Sequence 48, Appli
79	114	1.3	1299	4	US-08-954-740-48	Sequence 48, Appli
80	114	1.3	1220	4	US-09-489-039A-12113	Sequence 12113, A
81	114	1.3	2480	4	US-09-335-011-1	Sequence 1, Appli
82	114	1.3	2311	3	US-08-934-386-9	Sequence 9, Appli
83	113.5	1.3	3057	4	US-09-360-416-3	Sequence 3, Appli
84	112.5	1.2	572	4	US-09-543-681A-8138	Sequence 8138, Ap
85	112.5	1.2	949	4	US-09-538-092-1290	Sequence 1290, Ap
86	112.5	1.2	1403	4	US-09-705-872-1	Sequence 87, Appli
87	112.5	1.2	2108	4	US-09-538-092-87	Sequence 87, Appli
88	112	1.2	15281	2	US-08-471-119A-2	Sequence 2, Appli
89	111.5	1.2	1286	3	US-09-268-140-3	Sequence 3, Appli
90	111.5	1.2	2349	4	US-09-538-092-914	Sequence 914, App
91	111.5	1.2	2644	4	US-09-028-047C-2	Sequence 2, Appli
92	111.5	1.2	3056	4	US-08-508-836A-8	Sequence 8, Appli
93	111.5	1.2	3056	2	US-08-629-001A-3	Sequence 3, Appli
94	111.5	1.2	3056	2	US-08-874-266-2	Sequence 2, Appli
95	111.5	1.2	3056	3	US-08-644-274D-3	Sequence 3, Appli
96	111.5	1.2	3056	3	US-08-952-127-3	Sequence 3, Appli
97	111.5	1.2	3056	3	US-08-952-014C-3	Sequence 3, Appli
98	111.5	1.2	3056	4	US-08-984-090-2	Sequence 2, Appli
99	111	1.2	862	4	US-09-665-479A-14	Sequence 14, Appli
100	111	1.2	1206	4	US-09-245-928A-19	Sequence 19, Appli



```

101 111 1.2 1243 4 US-09-245-928A-17 Sequence 17, Appl
102 111 1.2 1247 2 US-08-472-93A-10 Sequence 10, Appl
103 111 1.2 1247 2 US-08-323-460A-10 Sequence 10, Appl
104 111 1.2 1247 2 US-08-461-146C-10 Sequence 10, Appl
105 111 1.2 1257 3 US-08-461-145C-10 Sequence 10, Appl
106 111 1.2 1257 3 US-08-947-823-5 Sequence 5, Appl
107 111 1.2 1257 3 US-09-245-928A-18 Sequence 18, Appl
108 111 1.2 1597 3 US-09-423-890-13 Sequence 13, Appl
109 111 1.2 1597 3 US-08-628-823-14 Sequence 14, Appl
110 111 1.2 2008 4 US-09-270-767-43754 Sequence 43754, A
111 111 1.2 2008 4 US-09-270-767-46774 Sequence 46774, A
112 111 1.2 804 4 US-09-526-193A-1 Sequence 1, Appl
113 110.5 1.2 2261 4 US-09-526-193A-1 Sequence 1, Appl
114 110.5 1.2 3144 1 US-08-246-982A-6 Sequence 6, Appl
115 110.5 1.2 3144 1 US-08-453-265-6 Sequence 6, Appl
116 110.5 1.2 3144 2 US-08-457-273B-42 Sequence 42, Appl
117 110.5 1.2 3144 3 US-08-556-419-21 Sequence 21, Appl
118 110.5 1.2 3144 3 US-09-041-886-15 Sequence 15, Appl
119 110.5 1.2 3144 4 US-09-538-092-1118 Sequence 1118, Ap
120 110 1.2 1575 4 US-09-917-254-83 Sequence 83, Appl
121 110 1.2 1867 4 US-09-824-574-5 Sequence 5, Appl
122 109.5 1.2 1055 3 US-09-031-563-27 Sequence 27, Appl
123 109.5 1.2 1055 4 US-09-392-277-27 Sequence 27, Appl
124 109.5 1.2 1055 4 US-09-258-000-27 Sequence 27, Appl
125 109.5 1.2 1315 3 US-09-031-563-25 Sequence 25, Appl
126 109.5 1.2 1315 3 US-09-031-563-25 Sequence 25, Appl
127 109.5 1.2 1315 4 US-09-293-505-10 Sequence 10, Appl
128 109.5 1.2 1315 4 US-09-392-277-25 Sequence 25, Appl
129 109.5 1.2 1315 4 US-09-392-277-25 Sequence 25, Appl
130 109.5 1.2 1315 4 US-09-258-000-25 Sequence 25, Appl
131 109.5 1.2 1315 4 US-09-060-939A-10 Sequence 10, Appl
132 109.5 1.2 2662 4 US-09-595-684B-31 Sequence 31, Appl
133 109.5 1.2 2662 4 US-09-046-572-3 Sequence 3, Appl
134 109 1.2 553 4 US-09-538-092-1252 Sequence 1252, Ap
135 109 1.2 2663 4 US-09-914-259-12 Sequence 12, Appl
136 108.5 1.2 1087 4 US-09-543-681A-5514 Sequence 5514, Ap
137 108.5 1.2 1186 4 US-09-328-352-5163 Sequence 5163, Ap
138 108.5 1.2 3118 2 US-08-457-273B-8 Sequence 8, Appl
139 108 1.2 778 5 PCT-US93-03076-3 Sequence 3, Appl
140 108 1.2 1513 5 PCT-US93-03076-2 Sequence 2, Appl
141 108 1.2 2482 4 US-09-252-991A-16967 Sequence 16967, A
142 107.5 1.2 1572 4 US-09-710-279-2906 Sequence 2906, Ap
143 107 1.2 335 2 US-08-875-062-3 Sequence 3, Appl
144 106.5 1.2 1232 4 US-09-248-796A-14563 Sequence 14563, A
145 106.5 1.2 3174 2 US-08-477-451-3 Sequence 3, Appl
146 106.5 1.2 1051 4 US-09-543-681A-8143 Sequence 8143, Ap
147 106 1.2 1068 4 US-09-248-796A-16119 Sequence 16119, A
148 106 1.2 1530 4 US-09-976-594-736 Sequence 736, Ap
149 106 1.2 1530 4 US-09-976-594-736 Sequence 736, Ap
150 106 1.2 1757 4 US-09-724-126A-15 Sequence 15, Appl

```

## ALIGNMENTS

```

RESULT 1
US-09-248-796A-19110
; Sequence 19110, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Ketch Weststock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248, 796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074, 725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096, 409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19110
; LENGTH: 1131

```

```

; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19110
Query Match
Best Local Similarity 18.5%; Pred. No. 2,4e-11;
Matches 222; Conservative 230; Mismatches 441; Indels 308; Gaps 60;

2.4%; Score 216; DB 4; Length 1131;

19 ELMTLLGRSALRE-----LSQIEAEINKMRRLLEGSLYKPPSSAER 64
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
52 QYTFPDNALSLRNCCDPYALADBEFLSSKEVLN-----ESPFLTK 96
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
65 YKANQDVASPLKEHGR--ISKFLGDEBSYQLQCYLOEDYRGTFDSYKTVQDR-- 120
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
97 DKQSPNIQT--RDISLKGIIYTNISVDISDGKLSNMLND--VLETRVICQINKKI 151
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
121 -----QSALILIKIAD-YYE-----ERTCILRCVHLITFYODERAPYR 159
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
152 PCKTAPQLEAIKSKLHKDKKYENKRLRLYSKILRRRIILIKIVLEL---NKSNSYA 208
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
160 VEVADCVDKLEKELVSKYRQFEEL-----YKTEAPYETHGNIMTERQV-SRMFVQ 210
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
209 ---SSSIQNGKIFLS-NQYLESLIVYIGNASHIMKSSYITGINKIEIDETIHNETVL 264
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
211 CLREOSMLRITFLYYAFEMAPSDLLVLTQKMEQGFGRQTRRHVDESTMPPYVRIG 270
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
265 CIEACKVLEL-----SVQNAVDAQVHLMFKLM---RDTNYSV---ALGPYV---S 308
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
271 YFSALILVEGMDI-----ESLHKCALDRREIHOPADGL----- 305
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
309 YHBAFSLIQLEFVLTQYLNNSFPSVNETLSSVMAD-VQVFYKVDHAIANPANNIF 367
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
306 -IQDDMCML-----TGGDIPHAIPVLIAMAL--LRHTNPEETSSVVRKIGTAIQ 355
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
368 VVYSGSIIILRKPYLOEYPELNESEKFLSQFNLSQLKRTIN-----LVNKKDN 418
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
356 LNFQYVLTIRLQSLASGANDCTTACMCYVG--LISFVLTSLHLTAGNOQDIPTACE 413
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
419 LDVFSSLKLNELKNDK-----LYSALITLIIASIPLTITSEV-----TSCI 463
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
414 VLADPSLP---ELFNGTEPTSGIILDSVCGMPFHLISPLIQLIRALVSGSTAKRY 469
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
464 LSVIGCPNNVVISFENNATQNAIIRT---KFLILSPYIQV--ASINGNFA---LH 515
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
470 SFDDKSPFVNELYKHPHDVISHEDGTWRORPKL--YR---LGGQNTLRIPQSTVQ 524
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
516 EFDULKSIVQVFKKEPNNNYQIDQNTBELVKTFTIDYPPPEANKKLSVLSGTAK 575
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
525 VM--LDDRAVLVWEXYSYSWTLFTCEIEMLVHVSADVTOHQCRVKPIIDLWHKV--I 580
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
576 ILPSANPDEVLTPLVYNGMAFLGRVLQVNSKIFNNSD---SETMELVITNLSLNLY 631
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
581 STDLSIADCLIPR-----TSRIYMLQRLTT--VISPPUVIASCNCLTVLAARNP 630
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
632 VIDNGVDISKVMEANSAYTDSIDILEVITRLLEQGLHRLNRYVLAVSNVLTNLNPFIS 691
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
631 AKVWTDLRHTGFLPFAHPPVSSIQMISAEGMAGAGYGNILMNSBOPOSGYTIAPLR 690
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
692 YRW-----PYLKK--SALFAQNGKEGLAAVIFGSI-----EMVNDYNTVLSLKL 736
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
691 ITTLVKGQSGSTQ-----SQGLVPCWVY--LKEMLPSYHKRYSNGVREIGICILEL 743
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
737 AEALIONCLSLQDYPBEKESVY--MLRFVGHVLDLPETFLYCRYNBOCQKLEIGVLLDT 795
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
744 IHALIWLCHETD-----LHSHTPSLQFCISGATYTAG 778
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
796 FSTILLASVYGIDEGVPANTKTKVFPADAASRLIDSEFLSDEDSPPAARLIT----- 846
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
779 QTVYINIGIGVDITDMVMAOPRSDGABGQGGQLIKTKVLAFSYVNNVIRLRKPSNVY 838
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
847 -TMIEMLSGDDLDLVELTDSSTFWYD-----NMHICALSSRLIITIR--TSEHL 892
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
839 SPL--EQALSQGHGNNLIIVLAKYIYKHDPAIPRAIQLIKRLAT-----VAPMSV 890
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 893 KPLAEKSLF---TKTSDLVLSRPFESVRKD-----ILDLLTLVTSGDMPDGTAP--SL 942  
Qy 881 YACLGND-AAAIRDAFLRLQSKIDMKIKMILEFLVAVV--TOPGLIEFL----- 941  
Db 943 LSHLSENAQVILHSLAADLNDSPDYKIKSLIYDFICAVMEGKEGLAVLFIITGRDVF 1002  
Qy 942 NLEVVDGSDGSKESFSGMWSCLHVLIELIDSGQDQRYMCP--PLHRA-AIAFLHALMD 998  
Db 1003 DYTSDQDTREATKLSL-----LQILKKIRIDMKRYENSVSIIHLYDALALACNSMTT 1053  
Qy 999 RRDAMLVLRTPKPKFEMNLTSPFLGTLSPSETSEPSI-----LETCALIMKILCLEIYY 1053  
Db 1054 AKES-----EHDDEFIQTILIGRVKQILDPDSSSEPSISRCVELKLVSKIALIILVLYFT 1108  
Qy 1054 V 1054  
Db 1109 V 1109

## RESULT 2

US-09-150-867-1  
Sequence 1, Application US/09150867  
Patent No. 6645748  
GENERAL INFORMATION:  
APPLICANT: Wood, Kenneth W.  
APPLICANT: Sakowicz, Roman  
APPLICANT: Goldstein, Lawrence S.B.  
APPLICANT: Cleveland, Don W.  
TITLE OF INVENTION: The Regents of the University of California  
TITLE OF INVENTION: Plus End-directed Microtubule Motor Required for  
FILE REFERENCE: 18557C-000110US  
CURRENT APPLICATION NUMBER: US/09/150,867  
CURRENT FILING DATE: 1998-09-10  
EARLIER APPLICATION NUMBER: US 60/058,645  
EARLIER FILING DATE: 1997-09-11  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 2954  
TYPE: PRT  
ORGANISM: Xenopus sp.  
FEATURE:  
OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)  
OTHER INFORMATION: member of the kinesin superfamily of microtubule  
OTHER INFORMATION: motor proteins  
FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (1) ..(472)  
OTHER INFORMATION: kinesin like motor domain  
FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (473) ..(2752)  
OTHER INFORMATION: rod domain  
FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (2753) ..(2954)  
OTHER INFORMATION: tail domain  
US-09-150-867-1

Query Match 1.6%; Score 147; DB 4; Length 2954;

Best Local Similarity 17.1%; Pred. No. 0.0012; Indels 508; Gaps 57;  
Matches 244; Conservative 202; Mismatches 474;

Qy 557 VSTADVIOHCORVKEPI-----DLVHKVISTDLSTADCLLP-----TSRI 597  
Db 1 MSEGAVAVVCRVRRLIQREGDQANLQWKAGNNTISQVDGKSNFDRVFNHSHSTQI 60  
Qy 558 YMLLORLTTVISPPVDVJASCV---NCLTVLAANPAKVTDLRHGTGLFPVVAHPVSSL 653  
Db 61 YQEI-----AVPIIRSLAQYNGTIFAYGOTSCKTYMTMGTPNSLIGIIPQAIQBV 111

Qy 654 SOMTSAEGMAGYGNILMNSQPGEGYVTAFLITLTVKQGL----- 699  
Db 112 FKII-----OEIPREFLAVSYWEIYNEYKOLLCDRRKKPLETRED 155  
Qy 700 -----GSTQSGVLP--CVMFVLK-EMLPYHKMRYSNGVREQIGCLILEIHAIL 748  
Db 156 FRRNVYVADLTIELVMVBEHYIOWIKKEGRKHVGETKNDHSRS-----HTIF 205  
Qy 749 NLCHETDLHSHTPSLOPLCTISLAYTEAGQVINIMIGVDTIDMVAAPRSDGAGQ 808  
Db 206 RPIVESRRNDPTNSEN-----CDGA-----VMVSHNLVDLASERSAQGAGG- 250  
Qy 809 GQGOLLITKVLAEFVNTNVRILKPSN-----VSPLEQALSGHAGNNLIJAVLAKY 862  
Db 251 -----VRLKEGCINNSFLIGQVITKLSGQAG-----F 281  
Qy 863 IYHKDPALPRLAIQQLR-----LATVAPMSYVACT-----GNDAAIR----- 902  
Db 282 INYR-DSKLTFRILQNSLGNKATVICTITVSPFETLSTIQFASHTAGVNTPHVNEVL 340  
Qy 903 --DAFLTLQSKIEDMRKMKMILEFLVAVETOPGLIEFLNLEVKDSDGSKESLGMW 960  
Db 341 DDEALLKRYRKEILDLKKOLENLE-----SSSETKAQAMAKEBHTQ--- 381  
Qy 961 SCLHVLIELIDSGQDQRYMCPPLHRAAIAFLHALMDRRDSAMLVLRTPKPKFME----- 1015  
Db 382 --LAEITQQLKEHEDRIW---HLTNIVYASSQESQDQR-----VKKRKYTAAPKIQ 431  
Qy 1016 -----NLTSPLFGLTSPSETSE-PSILETCALIMKILCLE----- 1050  
Db 432 NSLHSGVSDPDMLSRLPGNFSKKAKFSDMSPFE-----IDSVCTEFSDFDALSMDS 487  
Qy 1051 -----IYVVGSDLSQSLD-----TLK 1068  
Db 488 NGIDAEMNLASKYTHREKTSIHQSMIDGQISDSVQFHSSKENQOYLPRDQGMABCR 547  
Qy 1069 KFSIEKRPAYW-----SGVYKSLAVHVAETEGSSCTSLLEYQMLVSAMRMLLI 1117  
Db 548 KASFEKETSLOQLOQSKSEBEKELVQSFELKIALEBQLSVKAKNLEMVNS--REHSIN 606  
Qy 1118 ATTHADIMHLDTSVVRQLFLDVLDTGTALL--LVPAVNCILRLGSM--KCTLLILLRLQ 1173  
Db 607 AEVQTDV--EKEVVRKEMSVLIGDSGYNASDLDQDSVDDGRKLSSHDECIEHRKMLEQ 663  
Qy 1174 WKRELSYDELIGPITELLEGVLQADQ-----QLMETKAKKAFSAFIVLQKEMKV 1225  
Db 664 --KIVDLEEFIEMLNKXSNDKQKSSBQDMESIQLCEALMAEKANALBELALMRD--- 717  
Qy 1226 SDIPQSOVLNVCETLQOEVIYALFDQTRHSIALGASATEDKSDMETDCCSRSHRDQDG 1285  
Db 718 ----NFDNIILIE-NETLKRBIADL-----ERSL-----KENDTNEFELIEKETOKEH 760  
Qy 1286 VCVIGLHLAKELCEVEDDQSMWLQ-VTRRLPILPTLLTLEVSL-----RMKONLHFTBAT 1340  
Db 761 EAQL-IHEIGSLKKLVENAVENYQNLBEDLFTKTLKEGHIQLAELKRADNLOKVRN 819  
Qy 1341 LHLITLARTOOGATAVAGAGITQISICLPILSVYOLSTNGTAQPSASR----- 1389  
Db 820 FDLVSVM-----GDSEKLCBEITQLKQSLSDAAYVTRDAQKESFILSENILE 866  
Qy 1390 ---KSLDAPSWPQVYRLSMSIMEOLLKTLRYNF----- 1419  
Db 867 LKEKHEEDISNMYNOKERKASLFEKQLETKSNYKMEADLOKELOSARNEYINYLGLLAG 926  
Qy 1420 -----LPEALDFVGHQERTLOCLINAVRTV--QSLACL----- 1450  
Db 927 KVPRLLSRVELEKKVSEFSKQLEKALEBKQA--LENEVTCISEYFLFNEVECLKNQISK 985  
Qy 1451 -----EADHTVGTIQLQSNPKEMKHHQLQKMDIQVNTAGYLCQACTSLHSKRM 1501  
Db 986 ASEBIMLKQEGEHSASIIISKOEIIMQSBQIQLQTLDEVHTTOSKVOQTEBOYLEMKM 1045  
Qy 1502 ---LQHTLQNK-----NODGLPSAVVAVQVRPSPAASAPSSKQAPADTEAS 1546



```

Oy      1270  -TDDCSSRRRDQDGVCGVLAKEICEVDEGDSWL--QVRRRLPIPLTLTLEY 1326
           : : : : : : : : : : : : : : : : : : : : : : : :
Db      1147  DFTDASIIH-----PIVRLTDDSPELRSTAMDTLSSLVFQLEKKYQITIPMNKVL 12000
           : : : : : : : : : : : : : : : : : : : : : : : :
Oy      1327  SLRKONLH-----FT-----EATLHLLTLTARTOQATVAGAGITOSCPL 13717
           : : : : : : : : : : : : : : : : : : : : : : : :
Db      1201  RHRINQYVDVLCIRIVKGYTLADEEDPLIYQHMLASSGDDALASGPVEG----PMK 12566
           : : : : : : : : : : : : : : : : : : : : : : : :
Oy      1372  SVYQLSTNGTQTPSPASKSLDAPSW-PCVYRLNSLM-EQLKTLR-----YNLP 14211
           : : : : : : : : : : : : : : : : : : : : : : : :
Db      1257  KLHVSTINLQKAWGARRVSKY--DWLEWLRSLDELKIDSSPSPLRSCMALAQAYN--P 1312
           : : : : : : : : : : : : : : : : : : : : : : : :
Oy      1422  EALD-----PYGV-----HOERTLOCINAVRTYOSLACEADHTWGFILOSNFME 1469
           : : : : : : : : : : : : : : : : : : : : : : : :
Db      1313  MARDLFNAFSCMSSELNEDQODELIRSIELALTSODIA--EYQT--LINTAEFM-E 13655
           : : : : : : : : : : : : : : : : : : : : : : : :
Oy      1470  WHFHLPOLMRDIQ--VNLG--YLCQACTSLHSRKMLOHYLONKNGDGLPSAVA--QRV 15222
           : : : : : : : : : : : : : : : : : : : : : : : :
Db      1366  HSDKGPRLRDONGVLLGEBRAKCRAYAKLHYEL--EPQKPTPALISLISINKL 14223
           : : : : : : : : : : : : : : : : : : : : : : : :
Oy      1523  QRPSASAPSSSKQ-PAADTEASEQALH 1552
           : : : : : : : : : : : : : : : : : : : : : : : :
Db      1424  QQPEASGVLEAYAMGFGELEIQTATWYKELH 1454
           : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
US-08-305-790B-2
Sequence 2, Application US/08305790B
Patent No. 6492106
GENERAL INFORMATION:
APPLICANT: SABATINI, DAVID M.
APPLICANT: ERDJUMENT-BROMAGE, HEDIYE
APPLICANT: LUI, MARY
APPLICANT: TEMEST, PAUL
APPLICANT: SNYDER, SOLOMON H.
TITLE OF INVENTION: MAMMALIAN PROTEINS THAT BIND TO FKBP12
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: BANNER & ALLEGRETTI, LTD
STREET: 1001 G STREET, N.W., 11TH FLOOR
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20001-4597
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/305,790B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,967
FILING DATE: 27-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: KAGAN, SARAH A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107,47225
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 197430 BBWB UT
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2549 amino acids
TYPE: amino acid
TOPOLOGY: 1linear
MOLECULE TYPE: Protein
HYPOTHETICAL: YES

```

```

; ; ANTI-SENSE: NO
; ; ORIGINAL SOURCE:
; ; ORGANISM: Rattus rattus
; ; TISSUE TYPE: pheochromocytoma
; ; CELL TYPE: PC12
; US-08-305-7908-2

Query Match      1.6%; Score 146; DB 4; Length 2549;
Best Local Similarity 18.6%; Pred. No. 0.0012;
Matches 273; Conservative 246; Mismatches 564; Indels 388; Gaps 72;

QY    LVLTWFMFEGGFGSGKOTNRHLVDSTMPDF-----VDRIGFSALILVEGDIE 284
DB    LILTTPREPKMKQKPQWTMRTHFEAEKGFDETLAKEMNDDXI-HGALLINELVRIS 275
QY    SLHKCALDDERELHQFADGLI-----CODMDCLMLTEGDIPIHNAIPVLAMALLRHTNP 340
DB    SMEGBRL--REMMEIRITQQQLVHNKYCKD-----LMGGTKPRHITPTFSF----QAQPQ 325
QY    ETSSVVRKIGGTAIQIANVFQYLTRYLLQSLASGNDCTTSTAQCVCYGILLSTVLSLIAT 400
DB    QSNALVGILGYSHQ-----GLMGFASPSPFKSTLVBSRCRDIM----- 366
QY    LGNOODITDTACEVLADBSLELPFWGTERPESGLGIIDSVCGMPPHILPLLOLRALVS 460
DB    EEKFQVQC-----WVKCRSSKNSLIQT-----LTLNLPRIVA 401
QY    GKSTAKKVYSFLDKMSFYNELYKAHPDIVISHEDGTLMRQTPKLPLYLGQTNLRPEOG 520
DB    FRPSA-----FTD-TQYIQDTMNVLSGVKEKERITAFOALGILLSAVRSEFKVYLP-R 454
QY    TVGQVMID-DRAVIVRMESYS-----SWLFTCEIEMLLHVASTADIQHOCRVKPI 572
DB    VLDIIIRAALPPKQFAHROKTQVDAIVFTFC-ISMLARAMPG-IQ-----QDI 501
QY    IDLVHAYISTDLSTADCLRPITSIKYMLQTLTVISPPUNVIASCNCILVLAARNAPK 632
DB    KELLEPMILAVGLSPA-----LTAVLYDISROI POL-KKOIQGLKMISVLMMKP- 551
QY    VWTDLRHGTFPFVAHPVS--LSQMISABQM-----NAGGY-GNLL----- 671
DB    LRHGMMKGLAHOLASPGTLTPRASAIVASITLALRTLSFFBFGHSILOPVHCA 607
QY    NNSQPOGEYGVTAIFLRITT--LVXGOLSTOSQGLVPQMFVLEKML----- 719
DB    PSYHKRRNVSHGVREIOGLILELHAILANLCHEPDLSHSHTPISLOFLCISLAYTEA 777
QY    PDPRD--RY-----CV-----LASDERPAMHAQOENIOL-----FYAL 700
QY    GQTYININGIGVDITIDWMAOPR-----SDAGGGQGQGLIKTYKLAHSV 824
DB    NDQVFEIELICTVGRGISMNPAFWMPFLKMLIQITTELHSGIGRIKQSARMGLH 760
QY    TNNAVRLRKPSNNVSPLEQAL-----SQGHAGNNILAVIAKYIYKHDPALPRLA 875
DB    VSNAPRILRF--WMEPILKAILLKOPDDPPNGVINNVLATIGE----- 804
QY    IQLKRALTVAWSVAVACLGDAALIRD-----AFLTRLOSIEDMRIKYMILEFLYAV 930
DB    LAQVSGLERKKVDELFIYINDMLODSLSILAKRQVALWTLGQIVASTGYVERV 858
QY    ETOGELIELFNLEWKOSDGSKSFSGMSCLAVALLEISQOQDRWCPRPLHRAIA 990
DB    RKYPFLLEVLNLFKTEBQNSTREBAIRKVGGLGLL-----DPYHKNVIG 904
QY    FLHALMODRDSAMLVLTRKPKFENILTSPLFGTLSPSETSEPSLETICALINKI--IC 1048
DB    MID---QSRDASAV-----SLSBSKSQDSBDYSTSMELVNMGILP 942
QY    LEIYYVWGSI.DQSLKDTLKKFSLEKRPAYSGYVKSIAVHAATBESSCTSLELEYQMLV 1108

```

Db 943 LDEFYPAVSMV-----ALMRFPRDSLSHHHTVWVAITFIKSLGKCYQFLPQWN-- 994  
Qy 1109 SAMRMLLIATTHADIMHLTSVVRROFLDV-----LDGTALLIWPASV 1154  
Db 995 -----PFLNVIRVCDAIREPFLFOQGLMVSFVKSHIRPYMBEITVLMKEFWVM 1044  
Qy 1155 NCLRLGSMKCTLLILLRQWKRELGSVDEIIGPLTEILEGVLOAQOOLMEKTKAVPSAF 1214  
Db 1045 NT-----SLOSTIIL-LIEQIVVALLGGERLY--LPQLPHMLRV--FWHDSQGRIVS-- 1093  
Qy 1215 ITVLOMKEMKVSIDIPQYSQVLVNCETLOEVEIALFDOTRSL-----ALGSATEDKDSM 1269  
Db 1094 IKLLAIIOLFANLDDYLLHLL-----PRIVKLFDAPEVPLBPKKALETVRLTSL 1146  
Qy 1270 E-TDCCSRHRDQDGVCGVGLHAKELCEVDEBGDWL--QVTRRLPIPLTLITLEV 1326  
Db 1147 DFTDASRIIH-----DIVETLDOSPRLSTAMDTLSLVLQGLKQYIPIPWNVKLV 1200  
Qy 1327 SLRMRQNIH-----FT-----EATLHLLTLARTQOQAVAVAGITQISICPL 1371  
Db 1201 RHRINHQYVDLICKIVAGYTLADEEDPLIYQHRMLRSSQDALASGPVETG---PMK 1256  
Qy 1372 SVYQSTUNGTQATPSASKSLDAPSW-PGVTRLSNLM-EQLLTKLR-----YNFLP 1421  
Db 1257 KLVHSTIMLOKAWGAARVSKD--DWLEMLRLSLLEIKDSSPSLSKCMALAAQYN--P 1312  
Qy 1422 EALD-----FVGV-----HOERTLOCLNVRVQSLACEADHTVGTFLQSNMKE 1469  
Db 1313 MARDLFNAFVSCSELNEDQODELIRSEIALTSQDIA---EVTQT--LLNLAEFW-E 1365  
Qy 1470 WHFHLPOLMRDIO--VNLG---YLCOACTSLSHSKMLQHYLQNKNGDGLPSAVA--QRV 1522  
Db 1366 HSDKGPLRLRDNGVILLGERAACRAVAKALHYKEL--EFQKPTPALLESLSINNKL 1423  
Qy 1523 QRPSPASASAPSSSKQ-PADPTSESOQALH 1552  
Db 1424 QQPEASGVLEPYAMKFGELRIQATWEKHL 1454

RESULT 5

US-08-471-112A-3  
Sequence 3, Application US/08471112A  
Patent No. 6313264  
GENERAL INFORMATION:  
APPLICANT: Molnar-Kimber, Katherine L.  
APPLICANT: Faili, Amedeo F.  
APPLICANT: Caggiano, Thomas J.  
APPLICANT: Nakamishi, Koji  
APPLICANT: Chen, Yanqiu  
TITLE OF INVENTION: EFFECTOR PROTEINS OF RAPAMYCIN  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flanagan, Henderson, Faradow, Garrett &  
ADDRESS: Dunner, L.L.P.  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471.112A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/384,524  
FILING DATE: 13-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/312,023

FILING DATE: 26-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/207,975  
FILING DATE: 08-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Siekman, Michael T.  
REGISTRATION NUMBER: 36,276  
REFERENCE/DOCKET NUMBER: 01142.0058-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2549 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULAR TYPE: peptide  
US-08-471-112A-3

Query Match 1.6%; Score 144.5; DB 3; Length 2549;  
Best Local Similarly 18.6%; Pred. No. 0.0017;  
Matches 336; Conservative 238; Mismatches 628; Indels 605; Gaps 87;  
Qy 237 LVLTQMFKEQGFSGRQTNRLVDETMDF-----VDRIGFSALIIIVEGWDIE 284  
Db 217 LIITQREPKEMQKPOYRRTFEBAKGFDETLAKEKGNRRDRI-HGALLIINELVRIS 275  
Qy 285 SLHKCALDDBRELHOPADGLI-----CODMCLMTFGDIPHHAPVYLAALRHITLNP 340  
Db 276 SMGERL--REEMBEITQOQLVHDKYCKD---LMGFGTRPRITPTSP-----QAVQ 325  
Qy 341 ETSSVVRKIG-----GTA----- 353  
Db 326 QSNALVGLIGYSSHOGIMGFGTSPSPAKSTLYESRCRDLMEBKFDQVQWVLCRNSKN 385  
Qy 354 --TQNVFQYLTSL-----LQSLASGNDCTTS-----TACMCVGLLSFVLT 394  
Db 386 SLIQMTILNLLPRLAFAFPAFTDYOYLDTPMHALSCYKKEKERTAAPOALGLSVAYR 445  
Qy 395 SLELHTLGNQODIITDPA-----EVLADPSLPFLFWGTBPTSGIITLD----- 438  
Db 446 SEFKVYLPRLDITIRALPDKDPAHROKAMQVDAVFCISMLABMGIGIQDIEKLL 505  
Qy 439 ---SVCGMPSHLSPLOQLRALVSGSKTAKKYVSLDKMSFYNELYKRPYVISHDG 495  
Db 506 EPMWAGLSPALTAVALYDLSRQIPQK---KDIQDGLKK--LSLVLMHRP----- 551  
Qy 496 TLMRQTPK-LAYPLGGQTNLRIPQGT-VGOVWLDRAVYLVREVSYSWTLF----- 546  
Db 552 -LHHPGMPKGLAQGLASPGITLTPASDVGSITLART-LGSEFEGHSITQVRRICADH 609  
Qy 547 -----TCE--IEMLHVSTADVIQCHQRYKPIIDLVHXY-----STDL 584  
Db 610 FLNSEHKEIRMEARCTSRILTPSILHIGHAVVSQTVQVAVADVLSKLLVGVITDPDP 669  
Qy 585 SIADCLLPITSRILYM-----LQRLTVISPPVIVASCVNCLTV--LAANPA----- 631  
Db 670 DIRVCVASIDERFDHMLAQAENLOMLFVALNDQVEIRELAIC-TVGRISNNPFWMP 728  
Qy 632 -----KVTWTDLRHTGF-----LPFAVPSVLSQMSISAEQNNAGVGN----- 670  
Db 729 FLKMLIQITLEHSGIRIKESARMGLHVSNAFLRIIP-----YMEPILKALL 781  
Qy 671 LMSSEPOGEYGTIAFLRLITLVKVGQSGTQSGVPCVMEFLKEMPSYHKWRYNSH 730  
Db 782 KKDPPDPNPQYNNVATI-----GEL--AQVSGI-----EM----- 816  
Qy 731 GYREQIGCLILSLHAIILNLCHETDLSHSTSLQFL--CICSLAY----- 774  
Db 817 -----VDELFTIMDLQDSILAKROYALMTLQGLVASTGVVPEYRYKPYLLEV 867  
Qy 775 -----TEAGQ-----TWINING-----IGVDITDMWMAQPRSDGAGGCGQL 813

```

Db      668 LNFLEKTEQNGCTREARIVGLGALDPYKHKVNIQIDSRDASANSLSKSSQDSS 927
Qy      814 LIKTVKLAFSVTNNT-RLKPPSNVSP-----EQALSQGHAGNLIYLAKEYHK-- 866
Db      928 DYSTSEMILVNNGNPLDEFPYAVSNVVALMIRFDQSLSHHTTMMVQALTFKSGIKCV 987
Qy      867 -----HDPALPRLAIOQLKRLATAPNSVACLGNDAAIRDAFL--TR 908
Db      988 QPLPQWMTPLNIVRVCDAIREPLFOQLGMVSVKSHIRPYDEITVTLKREFWMTS 1047
Qy      909 LOSKIEDMRIRKMILEFTVAVETO-----PGLLEFLNLEVKDSDGS-----K 953
Db      1048 IQSTI-----ILILEQIVVALGGEFKYLPQLIPHMLRVFMHNSPGRIVSIKLAAIQ 1101
Qy      954 EFSIGMWSCHLAVL---ELIDSQOQDRKPCPLHRAPIALFHALMO--DRDSANLVL 1007
Db      1102 LFGANLDYHLHLLPPIVKLFDAPF-----APLPRKALLETVDRLTESLDTYASRII 1156
Qy      1008 RTKRFMENTLSPLEGTLSPPSESEPSILETCALIMK----- 1045
Db      1157 H-----PIVRLDQSPBELRSTAMDTLSSLVFLQKKYQIFIPMNKVLVRHRIN 1205
Qy      1046 -----IICLEIYVYVKG-SLDSQSLKDTL----- 1067
Db      1206 HQRYDVILC-----RIVKGYTLADEEDPLIYQHMLRSGGDALASGVENTGPKKXHLVS 1261
Qy      1068 -----KFSIEKRFAY--WSGVKSLAVHVAETGSS--SCTSLLE-YQML----- 1107
Db      1262 TINLOKAMGAARVSKODMELMRLSLLELKDSSPSLRSCWMLAQVYNMARDLFNAA 1321
Qy      1108 -VSAMRM-----LIATTHADIMHLDTSVVARQLEFVLDLDTKXLLLVPSVN 1155
Db      1322 FVSCWSELNEDQOELRSIELATLSODIAETOTLLMLAEFMESDGPPLDDNGIV 1381
Qy      1156 CLRLSGMKCTLLILLRQMKELGSDVDEILGPLETIEILGVLAQOQLEKTKAVFSAFI 1215
Db      1382 LIGRAKCAVAVAKALHKELEFQK-----GPTAILESLISINKKQPPRA---AGV 1432
Qy      1216 TVLQMK---EMKV-----SDIP-----QYSQ 1233
Db      1433 LEVAMKHFGELEIQATWYEKLEHWEDALVAYDKMDTKNDPBELMGLGMRCLGALGEGWQ 1492
Qy      1234 LVANVCE--TQOEVIAPFDOTRHSLAGSATE-----DKSMETDDCSRRHDOQPGV 1286
Db      1493 LHQOCCEWTL-----VNDSTQAMAMMAAANGLGQWDSMEBYTCMIP--BDTHDGA 1544
Qy      1287 ---CVLGLH-----LAKELCEVDED-----GDSMLQVTRL---PIPLTLTLYE 1325
Db      1545 FYRAVLALHQLFSLAQOCIDKARDLDAELTAMAGESYSRAYGAMVSCNHLSELEBYIQ 1604
Qy      1326 VSL-----BKKONLHTEATLHLLTLA-----RTQOATAVAG--- 1359
Db      1505 YKLVERREIRIQIWERLQCCORIVEDMOKILMVRSLVSPHEMDRMTLKYASLCGKSG 1664
Qy      1360 -AGITGSCILLSYUOLSTNGTAOTPSARSKSIDAPBWPQY--RLSISLMEQLKTLR- 1416
Db      1565 RLALAHKTLVLLLG-----DPS---RQLDHP--LFTVHPQVYTAAMKNMKWSARK 1710
Qy      1417 YNPLPEALDFVGVHQTLOCINAVRTVQSLACLEADHTGAF--TLQSLNFMK--EMWFIH 1473
Db      1711 IDAQOHMHFVOTMOQ---QAQHAIAIT-----EDQHKQELHMLMARCTLKGEMOYN 1760
Qy      1474 L-----POLMRDIQV-----NLGYLQACTSLSHSRKMLQHYLQNKND---GLPSA 1517
Db      1761 LOGINESTIPIKLYQYASATEHDSRWYKAMAMVAMFEAVLYHGYHQQADDEKKLHA 1820
Qy      1518 VAQRVQRPASASAPSSSKQPAADTEASFOALHT-----VOYGLKILSTTLAL 1569
Db      1821 SGANTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1878
Qy      1570 RHFTPDV 1576

```

```

Db      1879 MYTVPVAV 1885

RESULT 6
US-09-950-634-3
; Sequence 3, Application US/09950634
; Patent No. 6713607
GENERAL INFORMATION:
APPLICANT: Molnar-Kimber, Katherine L.
            Falliti, Amadeo F.
            Caggiano, Thomas J.
            Nakanishi, Koji
            Chen, Yangu
TITLE OF INVENTION: EFFECTOR PROTEINS OF RAPAMYCIN
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
            Dunner, L.L.P.
            STREET: 1300 I Street, N.W.
            CITY: Washington
            STATE: DC
            COUNTRY: USA
            ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/950, 634
FILING DATE: 13-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/471,112
FILING DATE: 06-JUN-1995
APPLICATION NUMBER: US 08/384,524
FILING DATE: 13-FEB-1995
APPLICATION NUMBER: US 08/312,023
FILING DATE: 26-SEP-1995
APPLICATION NUMBER: US 08/207,975
FILING DATE: 08-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Slekmann, Michael T.
REGISTRATION NUMBER: 36,276
REFERENCE/DOCKET NUMBER: 01142.0058-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4000
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2549 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-950-634-3

Query Match      1.6%; Score 144.5; DB 4; Length 2549;
Best Local Similarity 18.6%; Pred. No. 0.0017;
Matches 336; Conservative 238; Mismatches 628; Indels 605; Gaps 87;

Qy      237 LVLTMRKEGQFGSRQTRNRLVDTMPF-----VDNIGVFSALIVEGMDIE 284
Db      217 LILTTQREPKEMQRPQWRHTFEBAKGFETTLAKKGNRRDRI--HGALLILNELVRI 275
Qy      285 SLHKCALDDBREHLPQADGLI-----CODMDCLMLTFGDIPIHAPVILAMALLRHTINPE 340
Db      276 SMEGERL--KEEMELITQOOLVHDKYCKD---LMGFTKPRHITPTPSF---QAVQPO 325
Qy      341 ETSVVRKIG-----GTA----- 353
Db      326 QSNALVGLIGYSHQGLMGFTSPSPAKSTLVESRCRDLMEKFKPDQVCGVILCRNSKN 385

```

```

QY 354 --IOLNFOYLTRL-----LQSLASGNDCTTS---TACMCVYGLLSPLYLT 394
DB 386 SLIOWITLNLRLIAFAFRSAPFDPOYLODTYMHALSCYKKEKERTAAFOALGLLSVAVR 445
QY 395 SLEHLTLGNQODIIDTAC-----EVLADPELPFWTEPTSGLIITD----- 438
DB 446 SEFKVYLPFLVDIIRAPALPPKDFAHKQKAMQVDAITFCISMLARAMQPGIQODIKELL 505
QY 439 ---SVQMPHLLSPLOLRALVSGSKTAKKVYFIDKMSFYNNLHGHKPHVDVSHEDG 495
DB 506 EEMLAVGSPALTAVALYDSROIPOK--KDIQGLLKM--LSLVLMKRP----- 551
QY 496 TLMRQTPK-LIYPLGGQTNLRIPQGT-VGVYMLDDRAVLYWMEYSYSWTLF----- 546
DB 552 -LRHPGMPKGLAHQALASPEGLTTLPEASDVGSITLALRT-LGSFEFGHSILOFVHCAHD 609
QY 547 -----TCE--IEMLHVVSADVIOHCQVKEPIDLVHKVI-----STD 584
DB 610 FLNSENKEIRMEABATCSRLTSPSHLSGHAHVVSQTAOVVADVLSKLTVGCTDDP 669
QY 585 SIADCLLPITSRIVML-----LQRLTVISPPVDVASCNCLTV--LAARNA----- 631
DB 670 DIRVCVLASIDBERFAHIAQAENLOALFVALNDVFEIRBELAIC-TVGRLSMMPAFVMP 728
QY 632 -----KWTDLRHTGF-----LPEVAHPVSSLSQMSABEGNAGGYGN--L 670
DB 729 FLRKMILQILTELEHSGIGRIKEQSGARMGHLSNAPRLIRP-----YNEPLIKALIL 781
QY 671 LMNSRQGEVGVITAFILRLITTLVKQGLSGTOSGLVPCWFLKEMLPSTHKKRINSH 730
DB 782 KIKDBDPNPGVNNVATY-----GEL--AQVSGL-----EM--RKM----- 816
QY 731 GVREOIGCLIELHAHILNCHETDLHSHTPSLOFL--CICSLAY----- 774
DB 817 -----VDELFIIMDLQDSSLAKQVALMTLQGLVASTGYVVERPKKPTILLEV 867
QY 775 -----TAGQ-----TVINING-----IGVDTIMVMAAOPRSDABEQGQGL 813
DB 868 LINFLEKTEONQOTREARIVGLGALDPYKHKVNIWIDQSRDASAVLSSEKSDSS 927
QY 814 LIKTVKLAFSYTNVVI--RLKPPSNVVSPL-----BOALSOGHAGNNILAVLAKIYHK-- 866
DB 928 DYSISEMVMGNBLPDEFPYAVSVNVALMRIFROOSLSHHTVVOALTFPKSLGKCV 987
QY 867 -----HDPALPRLAIOILKRLATVAPMSVYACIGNDAAAIRDAFL--TR 908
DB 909 QFLPQVWPTFLNVRVCGAIRRELFQOLGMLVFSVXSHIRPYMDELVTTLMBREFWNNTS 1047
QY 1048 IQSTI-----ILLIEQIVVALGSEFKLYLFQOLLPHMLRFPMHNSPGRIVSIXLAIQ 1101
QY 954 EFSIGMNSCLHAVL-----ELIDSQODRWYCPRLHBAIAFLHALNQ--DRDSAMLV 1007
DB 1102 LFGANLDYHLHLBPYKLFDAPE-----APLSRKALETYVRLTESLDFDYASRII 1156
QY 1008 RTKREKMENTSPFLGTLSPSPSETSBSISIELCALINK----- 1045
DB 1157 H-----PIVRTLDSQSPELRSTAMDLSLAVQLGKQYQIFIPMNKVLVRIRIN 1205
QY 1046 -----IICLETIVVYVG-SLDSQSLKDTL----- 1067
DB 1206 HQRDVILIC--RIVKGYTLADEEBDPLLYOHRMLRSQGDALASGPVETGPKKLHVS 1261
QY 1068 -----KKSIEKREPAY--WSGYVSLAVHVAETEGS-----SCTSILE--YQML----- 1107
DB 1262 TINLQKAMGAARVSKXDMLEMLRRLSLLELKXSSSSSLASCMAAQAAYPMARDLFMA 1321
QY 1108 -VSAMRNL-----LIATTHADIMHLDVSVVRROFLDVIDGTALLLVASVYN 1155
DB 1322 FVSCWSELNDDODELIRISIELALTSODIAEVTOITLNLAEFWEHSDKGPLRLDNGIY 1381

```

```

QY 1156 CLRLGSMKCTILLILLBQMKRELGSVDEILGPLELIEGVLQADQOLMEKTKAVFSAFI 1215
DB 1382 LIGERRAKCRAYAKALHYKELEFOK-----GPPALLESLSISNNKLOQBEAA---Agv 1432
QY 1216 TVLQMK--EMKV-----SDIP-----OYSQ 1233
DB 1433 LEVAMKRGFELEIQTATYVEKLEHWEEDALVAVDKMDTNKDPBELMLQRMKCLMAGMWQ 1492
QY 1234 LVLANVC--TLQEEVIALFPQTRHSLAGSATE-----DYDSMETDDCSSRRHRDQDV 1286
DB 1493 LHQOCCEKMTL-----VNDETOAKARMAAAMAGLGQWDSMEEYICMIP--RDTIDGA 1544
QY 1287 ---CVLGLH-----IAKECEVED-----GDSMLQVTRRL--PILEPTLITLLE 1325
DB 1545 FYRANVLAHQDLHSLAQCCIDKARDLDAELTMAGESISRAYGAMVSCMLSELEEVIO 1604
QY 1326 VSL-----RMKONLHFEATLHLLTLA-----RTQOGATAVAG-- 1359
DB 1605 YKLVERREIIRQIMWERLQCGQRIVEDQKILMVSLLVSPHEDMRTMLKVASLOGKSG 1664
QY 1360 -AGITQICLPILSYVQLSTNGTAOTPSASRKSLDAPSWGVY-RLSMSLMEQLITLR- 1416
DB 1665 RLAAHKTLLVLLGV-----DPS--RQLDHP-LPTVAPQVYAYMKMMWSARK 1710
QY 1417 YNLEPEALDFGVGHQERTLOCLNAVRTVQSLACEADHTVGF-IILQSNFMK--EWHFH 1473
DB 1711 IDAFQNHQHVQMQO---QAQHAIAI-----EQQKHQELHKLMAFCFLKLGEOVLN 1760
QY 1474 L-----POLMRDIQV-----NIGYLQACTSLHSRMTQHYLQNRKGD--GLPISA 1517
DB 1761 LOGINESTIPIKVIQVYSAATEHDSRWYKAMHAWAVNPFVAILYKHQONARDEKKLRHA 1820
QY 1518 VAQVORPPSAASAAPSQKQAPADTEASBQALHT-----VOYGLKLITSLTAL 1569
DB 1821 SGANITNATTAATTAATTAATTAATSTEGSSESEASTENSPTSPBLQKVTEDISKTL-L 1878
QY 1570 RHFTPDV 1576
DB 1879 MYTVPAV 1885

```

```

RESULT 7
US-09-538-092-1112
/ Sequence 1112, Application US/09538092
/ Patent No. 6753314
/ GENERAL INFORMATION:
/ APPLICANT: Glot, Loic
/ APPLICANT: Mansfield, Traci A.
/ TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
/ FILE REFERENCE: 15966-542
/ CURRENT APPLICATION NUMBER: US/09/538, 092
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: 60/127,352
/ PRIOR FILING DATE: 1999-04-01
/ PRIOR APPLICATION NUMBER: 60/178,965
/ PRIOR FILING DATE: 2000-02-01
/ NUMBER OF SEQ ID NOS: 1387
/ SOFTWARE: Curaparseqformat Version 0.9
/ SEQ ID NO 1112
/ LENGTH: 2549
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc. feature
/ LOCATION: (0)..(0)
/ OTHER INFORMATION: Polypeptide Accession Number P42345
US-09-538-092-1112

```

```

Query Match 1.6%; Score 143; DB 4; Length 2549;
Best Local Similarity 17.6%; Pred. No. 0.0023;
Matches 262; Conservative 246; Mismatches 559; Indels 420; Gaps 65;
237 LVLTGKFEQKGRGKQTRNRLHVDETMDPF-----VDRIQFSALIVGMDIE 284

```



```

Db 217 LILTTGPEKMKQKQWRHTFEEBAKGFDETLAKKGMNRDDI-HGALLILNELVRIS 275
Qy 285 SLHKCALDDREELHQPADGLI-----CODMCLMTFGDIPHNAPVLAWALLRHTLNPE 340
Db 276 SMEBERL--REMEBEITQQOLVHDKYCKD-----LMGFETKRHTTPFTSF--QAVQPQ 325
Qy 341 ETSSVVRKIGTALQIANVOYLTRLRLSLASGNDCTTSTACMKCVYGLSFVLSLEHT 400
Db 326 QSNALVGLGYSSHQ-----GLMGF----- 345
Qy 401 LGNQODIIDTACEVLADPSLPFWGTEPTSGLGIIIDSV-----GMFPHLS----- 449
Db 346 -----GLTSPRAKSTIVSRCCRODMERKFOVCOWCYLKC 380
Qy 450 -----PLQLLRALVSGSTAKKAVSFLDKMSFYNELYKHPRDIVSHEDGILMR 499
Db 381 RNSKNSLIQMTILMLPRILAFRPSA-----FTD-TQYLDPTMNHVLSGVKKERETAA 434
Qy 500 RQTKLLYPLGGQTNRIPOGTGQWMD-DRAVLYRREYSYS-----SWLFTCEIE 551
Db 435 QALGLLSVAARSEKRYLPR-----VLDIIRALPRKPAKQOKAMQVDAVFTC-IS 487
Qy 552 MLHVTSTADVIQHCQVKPIIDVHKVISTDLSTADCLPITSRIYMLQRLTVVISPP 611
Db 488 MLAAMGBG--IQ-----QDIKELBERMLAVGLSPA-----LTVAVYLSQFIQOL--K 532
Qy 612 VDVIASCVCNCLTVLAARNPAKWTDLRHTGFLPVVAHFVSSLQWISAEKNAGCYGNLL 671
Db 533 KDIOGGLKMLSLVLMHKP-----LRHPGMPKGLAHQLASPGTLTPLEASDVG----- 580
Qy 672 MNSRQPGCEYVNTAFRLITTLVKQ-----LGSTOSQGLVCEWFLVKEM 718
Db 581 -----SITLALRTLSFEFEHSLTQFVRHCADHFLSHKEIRMAAARTCSLL 630
Qy 719 LPSYHKRNSHGVREQIGCLILELIALI-----NLCHETDLHSSHT 761
Db 631 TPSIHLSGHANVSGTAVQVADVLSGLVVGITDPDPDIRCYVLSLDRFAHLAQ 690
Qy 762 PSLQPLCISLAYTEAGQTVINIGIVDTIDMMAAQR-----SDGAEQ 808
Db 691 ENLOAL-----FYALNDQVEIRELALCTVGRLSMNPAPVMPFLRMILQITLEHS 744
Qy 809 GQGGLIKTVKLASVTNNVRIKRPSSVNSPLEAL-----SGHAGNNLAVL 859
Db 745 GIGRIKESQAMLGHLVSNAPRLRP--YMEPILKALLIKLKDPPDPENPGVINVLATI 802
Qy 860 AKYIYHKHPALPRLAIQILKRLATVAPMSVYACIGNDAAIRD-----AFLTRLOSXIE 914
Db 803 GE-----LAOVSLBMRKAVDELFIIMMLQDSSILAKQVALW 842
Qy 915 DMRIKVMLEFLVAVETQGLIELFLNLEVYKSDSGSKESLQMSCLAHLVLEIDSOQ 974
Db 843 TLGQLVASTGVVEPYRKYFTLLEVLNFLTQEQGRTREAIRVGLGLAL----- 894
Qy 975 QDRVWCPRLHRAIAFLHALMODRRDSAMLVLRKPKFMENTLSPLFGTSPSPSESEP 1034
Db 895 -----DEYKHKVNIQID-----QSRDASAV-----SLSESXSQDS 926
Qy 1035 SILEFCALIMKI--ICLEIYVVGSLDQSLKDLTKKSIKRPAYMVGYSKSLAVHAA 1092
Db 927 SDYISTEMLVMGNLPLDEFYPAVSMV-----ALMRFRDOSISHHTTVOAITTFIK 980
Qy 1093 TEGSSCTSLSEYQMLVSAMRMLLIATTHADIMHLDTSVVRQFLDV----- 1140
Db 981 SLGKCVQFLPQVM-----PTFLNVIRVCCDAIREPLFQOGMLVSFVKSIR 1028
Qy 1141 --LQTKALLVPASVNCRLSGMCKTLLILLRQMKELGSVDEILPRLIEBVGQA 1198
Db 1029 PYMEIVTLMEFWWMT-----SIOSTIIL-LIEIIVVALGGEFLGY--LQOLIPHMLRV 1081
Qy 1199 DQOMERTKAVFSAFITVLQMKEMKVDIPOYSOLVNVCEITLQEBEYALFDQTRSHL- 1257

```

```

Db 1082 --FMHNSPGRIVS--IKLLAAILFGANLDDVLLHLL-----PIYKFLDAPEARLP 1130
Qy 1258 ----ALGSATBEDXDSME--TDDCSRRKRDQDVCYGLHAKELCEVDEBDSWL--QV 1310
Db 1131 SRKALBTVDRLTESLDFTYASRIH-----PIVTLQOSPRLSTAMDTLSLVFOL 1184
Qy 1311 TRRLPIPLTLTLEVLAKMKNH-----F-----EATLHLLTLARTQOAT 1355
Db 1185 GKTYQITPMWVKVLYVHRIRHQRYDVLCRIYVGYTLABEEBPLLYQHRMLRSQGDA 1244
Qy 1356 AVAGAGITOSICPLPSVYOLSTNGTAQTPSARSKSLDAPSW-BGVYRLSMILM-EQLK 1413
Db 1245 LASGPVETG-----PMKHLVSTINLQKAMGAARVSKD--DWLEWMLRLSLLELKDSSP 1298
Qy 1414 TLR-----YNFLPEALD-----FYGV-----HOEBRLQCLNAVRYQSLACBEA 1453
Db 1299 SLRSCMALAQAYN--PARARDLFNAFVSCWSELNEDODELIRSEIALTSQDIA---EV 1353
Qy 1454 DHTVGFILQSNFPEKEMHFLPOLMRDIQ--VNLG---YLQOACTSLHSRKLQHYLON 1508
Db 1354 TQT--LNLAEFN--EHSDKGPLPLRDNGIVLIGBRAKCRAYAKALHYKEL--EFOKG 1407
Qy 1509 KNGDLPSSAVA--QRVQRPSPASAPSSSKQ--PAADTEASEQOALH 1552
Db 1408 PTALESLSLISNNKLOQPEAAGVLEVMKHFGELEFIQATWYKELH 1454

```

```

RESULT 8
PCT-US95-06722-12
Sequence 12, Application PC/RUS9506722
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Immunosuppressant Target Proteins
NUMBER OF SEQUENCES: 25
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06722
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/250,795
FILING DATE: 27-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/250,795
FILING DATE: 20-DEC-1994
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2549 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-06722-12

```

```

Query Match 1.6%; Score 143; DB 5; Length 2549;
Best Local Similarity 17.6%; Pred. No. 0.0023;
Matches 262; Conservative 246; Mismatches 559; Indels 420; Gaps 65;

Qy 237 LVLTLMKEQFGSGRQTRNRLVDTMDPF-----VDRIQFSAIILVEGMDIE 284
Db 217 LILTTGPEKMKQKQWRHTFEEBAKGFDETLAKKGMNRDDI-HGALLILNELVRIS 275
Qy 285 SLHKCALDDREELHQPADGLI-----CODMCLMTFGDIPHNAPVLAWALLRHTLNPE 340
Db 276 SMEBERL--REMEBEITQQOLVHDKYCKD-----LMGFETKRHTTPFTSF--QAVQPQ 325
Qy 341 ETSSVVRKIGTALQIANVOYLTRLRLSLASGNDCTTSTACMKCVYGLSFVLSLEHT 400
Db 326 QSNALVGLGYSSHQ-----GLMGF----- 345

```

```

QY 401 LGNODIDTACEVLADPSLPFWGTEPTSGIILDSVC-----GMFPHLS----- 449
DB 346 -----GTSBSPAKSTLVESRCRDMEEKFDQVCOWILKC 380
QY 450 -----PLLOLRALVSGSKTAKKYVSFLDKMSFYNELYKRPDVI SHEDGTLMR 499
DB 381 RNSKNSLIQMTILNLPRLAERPSA-----FTD--TOYLOPTMNVHVSCKVCKERTNAF 434
QY 500 RQTPRLPLPLOGQNTLRIPOGTGVQVMD--DRAYLVREHYSIS-----SWLFTCEIE 551
DB 435 QALGLSVAVSEFVYPR-----VLDIIRALPDKDFAHKQKAMQVATVFTC--IS 487
QY 552 MLHVSTADVIQHCORVKEPIIDLVHKTISTDLSTADCLLPITSRIYMLORLFTVISBP 611
DB 488 MLARMBG--IQ-----QDIKELLEPLAVGLSPA-----LTAVLYLSQIQOL---K 532
QY 612 VDVLASCVNCLTVLAARPAKWTDLRHTGFLPYAHFVSSLQMSIAGNNAGYGNL 671
DB 533 KDIDQGLKMLSLVYMHKP-----LRHPGMPKGLAHQLASPGLTTLPEASDVG----- 580
QY 672 MNSEQPOGEYGVTLAFLRLITTLVYQ-----LGSTQSQGLVPCWPFVLKEM 718
DB 581 -----SITLALRTLGSFEEHSGLTQFVRHCADHFLNBEHKEITRMAARTCSRL 630
QY 719 LPSYHKWRYNSHGVREOIGCLLELIALIL-----NLCHETDLHSHT 761
DB 631 TPSHILISGHAAVVSQZAVQVADVLSKLVAGITDPDPDIRCYVLSLDERFAHIAQA 690
QY 762 PSLOFLCISLAYEAGOTVINIMIGVDTIDMMAAPR-----SDGABEQ 808
DB 691 ENLOAL-----FALANDQVEIRELALCTVGRLSNNPAPVMBELRMLIQLITELHS 744
QY 809 GOGOLLIKTVKLAVSTNNVIRLKPSPNVSPLEQAL-----SOHGAHGNLAVL 859
DB 745 GIGRIKESAMGLHIVSNAPRLIR--YMERIILKALIKLKDPPDPNPNVINVATI 802
QY 860 AKTYHKNRDLPRLAIOQLKRLATVAPMSVYACLGNDAAIRD-----AFLTELQSKIE 914
DB 803 GE-----LAQVSGLEMRKWDLEFIIIMMLQDSILARQYALW 842
QY 915 DMRIKWMLEFLTVAVETQGLIELFLNLEVKDSDSKERSLGMSCLAHVLELDSQ 974
DB 843 TLGQIVASTGVVERPKYPTLLEVLNLFKTEONOGTRRAIRVGLGLAL----- 894
QY 975 QDRYWCPELHRAAIAFLHALMODRDSAMLVLRTEKFEWENTLSPFLGTSPSESEP 1034
DB 895 -----DPYKHKNVIGMID-----QSRDASAV-----SISEKSKSDS 926
QY 1035 SILETCALIMKI--ICEIYVVGKSLDOSLKTLLKFSIEKRPAYSGYKSLAHVAE 1092
DB 927 SDVSTSEMLVVMGNLPDEFYPAVSMV-----ALMRIIPDOSLSHHHTWVVOAITPFIK 980
QY 1093 TEGSSCTSLLEYOMLVASAMRMLLIATTHADIMHLLTSVVRQLEFLV----- 1140
DB 981 SLGKACVOFLPQVW-----PTFLANIVVCCGAIRHEFLFQOLGMLVSEFVSHIR 1028
QY 1141 --LDGTALLIVPASVNCRLGSKCTLLILLRQMRKELSGVDEILGPTTEILEGLQA 1198
DB 1029 PYMDEIYTLMEFVWVNT--SIOSTIIL--LIEQIVVALGSEPKY--LPQILPHMLRV 1081
QY 1199 DQOLMEKTRAKVPSAFTVLQMKEMKYSDIPOYSQVLNVCETIQOEIVALLPQTRHSL- 1257
DB 1082 --FHHDSFGRTVS--IKLLAALIQFANLDDRYHLL-----PRIVKLPAPPEARLP 1130
QY 1258 -----ALGASATEDKDSME--TDDCSRSRRDQRDGVCLGLHLAKELCVDSDGSMWL--QY 1310
DB 1131 SRKALETVORLRESDLFTDYASRIH-----PIVATLDSQPELRBTADTSSLVFOL 1184
QY 1311 TRRLPIPLTLTLEVSLRKNKONLH-----FT-----EATHLALLTLARTOGAT 1355
DB 1185 GKXKQIPIPMWNKVLVHRINHGRYDVLICRIYGVYTLADEEDPELIIYQHRMRSQGDA 1244
QY 1356 AVAGAGITQSIGLPLSLVQSLSTNGTQTSARSKSLDAPSW--PGYVRLSMSLM--EQLLK 1413

```

```

DB 1245 LASGPVETG---PMKLIHSTINLQKAWGAARVSD--DWLEWLRSLLEILKSSSP 1298
QY 1414 TLR-----YNLEPEALD-----FVGV-----HOERTLOCLNVRTVQSLACLEEA 1453
DB 1299 SLRSCWALQAQV--PWABRLPNAAPVSCWSELNEDQDLSIELALTSODIA---EV 1353
QY 1454 DRYVGFILQSLNPMKEMHFLPOLMRDIQ--VNLG---YLCQACTSLHSRKLQHYLON 1508
DB 1354 TOT---LNLIAEFM--EHSDDGFLPLRDNGI VILGERAACRAVAKAHYKEL--EFQKG 1407
QY 1509 KNGDGLPSAVA--QRYORPSPASAAAPSSGQ--PAADTEASEQALH 1552
DB 1408 PTRAILESLSINNKLOQPEAAAGVLEFANKHGELEIQTWYKELH 1454

RESULT 9
US-09-914-259-11
; Sequence 11, Application us/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-914-259-11

Query Match 1.6%; Score 140; DB 4; Length 3878;
Best Local Similarity 18.7%; Pred. No. 0.01;
Matches 307; Conservative 232; Mismatches 583; Indels 520; Gaps 70;

QY 59 PSSAEYKANKDVASPLKELGRISKFLGLDEBOSVOLLCYIODEYRGTROSKVTIAD 118
DB 2081 PISEHQTRVEQOLANLKEKTKSCSEL--LSKEQ-----LORDIQERNEIEKLEPR 2131
QY 119 EROSOALIKINDY--YEER-----TCILRCVLIHLTYPRODRHPRVYACVDBLKE 172
DB 2132 VRELEQALVSADTFQKVEDRKHFGAVEAKPELSLEVOLOAR-----DAIDRKEKE 2183
QY 173 LVSKYRQ--QF--BELYKTEAPTWETHGNLMTEROVSRWFVQCLREOSMLLEIIFLYAYF 229
DB 2184 ITVLEQQLQFQREBELNKNVEVOQLHMQLEIQKSTRTLOLEQENKLFK----- 2234
QY 230 EWAAPSLVLVTMFKFQGGSGRQNRHLVDEMDPVDRIQYFSALIVEGMDIESLH-- 287
DB 2235 ---DMERKGLAIKESDMSSTQ--DOHVL-----FGKFAQIIOKEVEIIDLNQ 2279
QY 288 -----KCALDDR--REIHOFAQDGLICODMCLMTFFGDIPIHAPVLAMALLRHTL 337
DB 2280 VTKLOQOOLKITTDNKNVIEKMLIRD--LEIQIECLM-----TSWN 2314
QY 338 NPEBTSSVVRKIGGTALQLVNFQVYLRLLQSLASGANDCTSTACWCYGLLSFVLTSL 397
DB 2315 --SDQCVKRNNEEIEQLN--EVIEKLQOEILANIGQK----- 2352
QY 398 LHTLGNQODIITDABVYLDAPSLPLFNGTEPTSGIILDSVCGMFPHLSPLIQLRA 457
DB 2353 AHSLSHADSLKHQLDVVAEKL--ALBQOVETANEBMTFMKVLKETNFKMQLTQELFS 2411
QY 458 LVSGSKTAKVVSF-----LDKQSFYNELYKRPKPH--DVISHEDGTLMRQTPGLALP 508
DB 2412 LKRRESVYEKIOSIPENSUNVAIDHLS-----KQKPEFLVLTDBA-----LKS 2455
QY 509 LGGQTNLRI--PQTVGOVMDLDRAYLVREYSSWTLFTCEIEMLLHVSTADVIQHC 566

```

```

Db      2456 LENQYFYSFEENGKSTINLETR-----LQLEST----- 2486
Qy      567 QAVKPIIDLKVKVITDLSIADCLPI-----TSRIYMLQRLTTVISPPVDVIA 616
Db      2487 -----VSAKDLELTQCYQIKQMOEGOFETEMLOKIKVNLQKIVEE--KVAA 2532
Qy      617 SCVNCLTVLAARNPAKWTDLRHGTPLPVAHPVBSLSQMSISAEQNGVGNLMLNSQ 676
Db      2533 ALVSGIQLEAVQEVYKFCOD-----NOTTISSEPERT-----NIONLNL 2571
Qy      677 POGEGVTIAPLRLITTLVKQ-----LSTQSQGLVPCVFMVLEKMLPSYHKWY 727
Db      2572 REDELGSDISALTNLISESQVEMWHTSLIEKQVEIAEKGVAEKEKKLELQKLEGG 2631
Qy      728 NSHGVR---QIGCILELILHAILMCHETDLSHSTPSLOPLCISLA-----YEA 777
Db      2632 NEKKQREKEKRSPODVEVLTKTTELPHSNE-ESGFNLELALRAESVATKALASYEK 2690
Qy      778 GOTVININGIGVDTIDMWAAQPRSDGALGQGGOLIKTVLAFSVTNVIRLKPSPV 837
Db      2691 AEKL-----QEBLVNE-----TN----- 2704
Qy      838 VSPLEALSQ---HGAGNNLJAVLAK-----YIHKDPAIPRLAIQULKELATVAPWS 889
Db      2705 MTSLOKDLQVARDHLAAEKESILIEKDETEVQSKKACWEPRLPYLKSISAGTGT 2764
Qy      880 VYACIGNDAAI---RDAFLTRLOSKIEDMRKIMILEFLIYAVETOPOLILFLNLEYK 946
Db      2765 LKISSNOTPOLIVNAGIQIMIOGECSEEBETEISOPTERIKEMQELHAAEILDMESR 2824
Qy      947 DGSOG---SKERSLGM-----NSCLHVLTELIDQOQORW--CPPLHRA-----AI 989
Db      2825 HISSETTKREHYAVQVLKEECGTAKAYIQCLRSKEVGFYNNCFSTLCSGSMGCI 2884
Qy      990 AFLHALMOD-----RRDSAMVLRTKPKFWMNLTSPFG-----TLSPPS----- 1029
Db      2885 YLTHSQGDIASEGGESESATDSFPKKIKGLLAHVNEHQVLSLSTEPSDDEDSHI 2944
Qy      1030 -ETSPSILETICALMKITCILEIYVVKSGSLDQSLKDTLKKSIEKRPAYSGVYKSLAV 1088
Db      2945 QOVSEPMLEERKAYINTI-----SSILKDLTKWQLORE----- 2977
Qy      1089 HVAEHESSCTSLIEYQMLVSAMWMLIATTHADIMHLDTSVVAROLFVLDGTAKLL 1148
Db      2978 --AEVYDS-----QSHESFSDMRGELLAL-----QOVFL-----ERSVL 3012
Qy      1149 LVPASVNCIRLGSMTCTLLILLRQMKRELGSVDEILGPLEILEGV--LOADQOLMEKTK 1207
Db      3013 LAAPTELJALGTDAVGLNLEORIQ-----QGEVQAMCEGLQKAD 3057
Qy      1208 AKVESAFITVL--QMKEMKV-----SDIPQSVLVNVCETLOEVIATLFDQRHSIA 1258
Db      3058 RRSLSSEIOAHQAMNGRKITLKRQESKEPQOELELYNI--QOKOSQMLEM--QVELISM 3114
Qy      1259 LGSATEDDSMETDCSRSRHRDQDGVCVGLHAKELCEVDEEDGDSMLQVTRRLPIIP 1318
Db      3115 KDRATELOEJUSSEK-----KVVAELKSELAQ----- 3141
Qy      1319 TLLTTLLEVSLRMK--ONLHTEA-----TLHLTL--TLARTQOGATAVAGAGITQ 1364
Db      3142 ---TLTELETLKQAHKHLKELEAFRLVEXDITDEVHLINDTLASBOQKSRRLQALERE 3198
Qy      1365 SICPLELVYQULSTNGTAQTPSARSKSLDAPSPGVYRLSMGLHQLKTLRYNLPAL 1424
Db      3199 KAKL-----GRSEERDKBELE-----DLKFSLESQOKRPNLOINLL----- 3233
Qy      1425 DFGVGHORBTLOCLN-AVRTVQSLACLEADHTVGFIILQSNFMKEMHNLPLQMRDIOV 1483
Db      3234 -----LEQOKLLESQOKIESQMLYDA-----QLSEEQR--NIELQVLTLESSEV 3278
Qy      1484 NLGYLQCACTSLHSRKMLOHYLQNKNGDGLPSAVALORVORPPSAASAASSKOPADT 1543

```

```

Db      3279 RI-----REMSSTLDRELEHAQLQSSDGTG-----QSRPP-----LPSEDL 3315
Qy      1544 EASEQALHTVOYGLKTLKST 1565
Db      3316 LKELQKOLEKGRSRIVELNET 3337

RESULT 10
US-09-270-767-44725
; Sequence 44725, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7325-094
; CURRENT APPLICATION NUMBER: US/09/270/767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 44725
; LENGTH: 919
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-44725

Query Match      1.5%; Score 136.5; DB 4; Length 919;
Beet Local Similarity 18.9%; Pred. No. 0.0015;
Matches 173; Conservative 134; Mismatches 322; Indels 287; Gaps 44;

Qy      944 EVKQSDSKERSFSLGMSCLHVLILDSQOQDRYWCPEPLHRAAIAFLH-----ALWQ 997
Db      30 EKQDKDKEQ-----PELSBDQ-----LQEBELVLVHQLEPDAKLXQ 70
Qy      998 DRDSAMVLRTKPKFWMNLTSP-----FGLTSP-----PSETSEPSILETICALMKI 1046
Db      71 PTLSEMAKLIRASTYMTSVKPELKMRYHETMKTYYKHMENQAR-----QLCADIIISV 126
Qy      1047 ICLIEIYVVKSGSLDQSLKDTLK--KFSIEK--RFAVWS--GYVYSL-----AVHVAEHESSC 1098
Db      127 LSWTN-----GSGKDCIAYRFLCDRKQIGMGHBYVHHLGSEISAHYHDTTGDR 177
Qy      1099 TSLIEYQMLVSAMW-----LTLIATTHADIMHLDTSVVAROLFVLDGTAKLL 1148
Db      178 VQLEILVQIIPYNNHNAHADCDLLI-----EIDHLLHSDPVDESAYPR----- 225
Qy      1149 LVPASVNCIRLGSMTCTLLILLRQMKRELGSVDE-----ILGPLEILE 1193
Db      226 -----CLYLOS--CYPVDPDNTIILETALQLSRKFNGYQAMRLALMINDMDKIGE 276
Qy      1194 GVLQADQOLMEKTKAKVSAFITVLQMKEMKVSIPQSVLVNVCETLOEVIATLFDQT 1253
Db      277 IFKEKEBPAMQQLAFMLARQOICLELDEL-----VPDDDL-----EIMSNANLN 323
Qy      1254 RBSLALGSATEDDSMETDCSRSRHRDQDGVCVGLHAKELCE----- 1299
Db      324 KHFLNLARELDIMEKTEBEDIYKSLHDSRSFASIQVDSAKQNLAAFPVNGFVNAFGCV 383
Qy      1300 ---VDEGDSMLQVTRRLPIIPT-----LITTEVSLRM--KONLHTEATL----- 1341
Db      384 DKLASBDDNKMLYKXKKEGMLSATASLGLIILMDVDGGLTMDKYLVSTDDINKGALIA 443
Qy      1342 -----HLLTLARTQOGA-----TAVVAGITQISICPLL--SYVOL 1376
Db      444 CGIYVNGIRNEVDPAHLLSDYIDQNSCMRVGAILGLIAYAGNSRSIVIDTLKTVSEF 503
Qy      1377 STNGTAQTPSARSKSLDAPSPGVYRLSMGL-----EQDLKTL----- 1415
Db      504 GSNMNS---SASVEL-----GITALSIGLIVSGSCNSEITEILLQIMGTAKDLMDT 554
Qy      1416 --RYNLPALDFFGVGHORBTLOCLNAYRTVQSLACLEADHTVGFIIL-----QLSNF 1466
Db      555 YTRFLFLGLGLIYLIG--RQKST-----EAV--MMTLEVELEEPYRSNATTVWDICAVAGTGV 607

```



Db 1948 TNEP---HLETFEESIOGF-----QRTTELEK-----LLE----- 1978  
Qy 1455 HTVGFILOLSNFMKMHNLPLQMLMDIOVNIQYLCOACTSLHSHKMLQHTLOKNGKGL 1514  
Db 1979 -----YMPFW-----LKNLKFCKSNDSDKCLKVSQILDK-----L 2009  
Qy 1515 PSAAQVORPPPSAASAPSSKOPADPEASEQALH-TVOYGL-----LKLSTLAAL 1569  
Db 2010 INLTIDQKEMPSVOAKIWGSIQIPELIDWLVNPLHKSITYGSGPQVEIMADTAYAL 2069  
Qy 1570 -----RHRTPDVCOIL-----LDQSL-----DLAEY-NFLPALSFTTPTPDS 1606  
Db 2070 ASANQVLSKQYITRICKVMDSCNTPVOYLEQHMMMDIALGRVYLMLSFNCL-----D 2126  
Qy 1607 VAPSEGTILATV-----NVALNMLGELDKKEPELTQAVGLSTOEGTPT 1650  
Db 2127 VATSPVYLFHTITFLVCSGSLMSRASTHGLVINIHSICTCTNP-----SFSEAOHV 2179  
Qy 1651 LK-SILMTMENCFLIL-----SQAMRYLRDPAVHPRDK-----QRMKQELSSLSLTL- 1698  
Db 2180 LRLSLDESLPK-FYLLPFGISKVSAVAATFRRSGCRHPTDKMLGNERVTOPLPADRELS 2238  
Qy 1699 LSSL 1702  
Db 2239 LPSTL 2242

RESULT 12  
US-09-510-791-1  
Sequence 1, Application US/09510791  
Patent No. 6365126  
GENERAL INFORMATION:  
APPLICANT: Zhong, Yi  
APPLICANT: Guo, Hui-Fu  
APPLICANT: Tong, JiaYuan  
TITLE OF INVENTION: Improvement of Learning and Short Term  
FILE REFERENCE: 1314, 1047-002 with Neurofibromatosis 1 (NF1) Expression  
CURRENT APPLICATION NUMBER: US/09/510,791  
CURRENT FILING DATE: 2000-02-23  
PRIOR APPLICATION NUMBER: US 09/046,745  
PRIOR FILING DATE: 1998-03-24  
PRIOR APPLICATION NUMBER: US 60/041,469  
PRIOR FILING DATE: 1997-03-24  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 2802  
TYPE: PRT  
ORGANISM: Drosophila  
US-09-510-791-1

Query Match 1.5%; Score 136; DB 3; Length 2802;  
Best Local Similarity 18.8%; Pred. No. 0.014;  
Matches 339; Conservative 244; Mismatches 635; Indels 586; Gaps 93;

Qy 169 LEKELVSKYRQOFEELYTEAFTW-ETHGNLTERQVSRWFOCLREOSMLLEIFLYYA 227  
Db 755 LOKRIMTLR-KIEHCVAGVQPAWETFRNWEVSSKVLQTYPKGGEQDQ----- 803  
Qy 228 YFENAPSLVLTKFKKQSGR-----QTRHLVDENMDPVRDRIQFSALILVEGDI 283  
Db 804 -----AEVF-HRGMGKRASHQSSHELEQINENANMTWPLLAGV----- 845  
Qy 284 ESLHKALDRRELHQFADGLICODMCLMLTFGDIPHHAFLVLAAMLRHTLTPETS 343  
Db 846 -CLHGRS-SRQMLQSQNN-----ASLG-----SLAQNSLXSSSTS 881  
Qy 344 SVKRIGSTAIQLN-----VFOYLTLLQSLASG-----G 373  
Db 882 SGHSGHSSTVSLTLPPAPPODVSYCPVQGLRLRLVCSNEKIGLTIQKVKELVG 941

Qy 374 NDCTSTACMCVYGL-----SFVLSTSLHTLGNQODIIDTA 411  
Db 942 EEMSTQ-----LYPLFPOVAIYKFPDQOQVNVNTDINTQIERTIYIMKSIDLPK 996  
Qy 412 CEVLADPSLPBLFMGTBPTSGIILDSVGMF-----PBL-----LSPLQLL 455  
Db 997 AN--KDPNNDQ-----PSPSEHLGV--TSIEGMGLIYRVHLDWTVYAIRIKTYLCQLV 1048  
Qy 456 PALVSGKSTAKKIVSFLDKMSFYNELYKHKHVDVSHEDGLMRQTKLXLYPLGQ--- 512  
Db 1049 EVMKMRGD-----LAFRQEMKFRNKLVLYL-----TDWVGTHQIAPSSADAA 1094  
Qy 513 --TNLRIPQGVGMMLDRAYLVWMEYSYSWTLFTCEIEMLHVSTADVIQRCQRYK 570  
Db 1095 ILTNTSLFPRDLQACMAVALLR-----GLPQEPESRGDLMD-----AK 1137  
Qy 571 PIIDVHKVISTDLSIADCLPITSRIYMLQRLTVVISPVDVYASVNCU---TVLAA 627  
Db 1138 SALFLKYFTLPMNL-LINDCISSEAKEM---NNTPLPRPMAAGLTLALRNATILAM 1193  
Qy 628 RN--PAKVTDLRHT--GFLPFV--AHVVSLSQMS-----AEGMAGGYNL- 670  
Db 1194 SNLGNANIDSGIMHSIDGYNPDLOTBAFMEVLTQIIQOGTEPDTLAEVLADEFQLV 1253  
Qy 671 -LANSQOPQGEYGVTAFLRLITTLVKGLSGTSGQGLVPCVMFYLKEML-PSYHKMRYN 728  
Db 1254 QLVTMISKXGELPIAMALANVYTTISQMDLAK-----LVYTLFPAKILSLBLMMMYR 1307  
Qy 729 SHGVREQICLILBLIHALIIMLCHETDLHSHSTPSLOFLCISLAYTEAGOTVINIMIG 788  
Db 1308 EWEVSD-----CMQT-LFRGNSLGSKIMARCFKY-----G 1337  
Qy 789 VDTIDMMAAPRSDGABGQGGOLLITVYKLASSYTNVNRVLRKPSNVVSELEALSQH 848  
Db 1338 ASYQMLLEPLIR-----PLDDEEBETCEV--DPAKLPTEI---ED----- 1376  
Qy 849 GAHGNMLIYAVLAKY---LYHGDPLPRLALO-----LAKRLATVAPMSVYACIGDA 898  
Db 1377 --HRRNLIALQKYPDAIINSSDRPPQJBSMCHLYOVLKSRPNL-----LQNNI 1426  
Qy 899 AAIRDAFLTRL-----OSKIEDMKIKMILEFLTVAVEFQGLIEL 939  
Db 1427 GAVGVYFLRFLINPAIVAPQELGIYDKOVHSAKRGIMLMSKILQNIANHVEFSKEQML 1486  
Qy 940 FLNLEVKDQSGSKSEFSLGMS-CIHAIVLELIDSGOQRYKCPRLHQAALFLH-AW- 996  
Db 1487 CFNDFLRHFAGRFFLIQIASDC-----ETVDQTSNM---SFISDANYLALHRLMT 1537  
Qy 997 -----QDRPSAMLVLRTPKPFMENTLSPLEGTSPSESEPSILETCALIMKI 1047  
Db 1538 HOEKIGDYLSRRDHKAVGRPRDKM-----ATLLAYLGPPBKRVDSHMFSS----- 1586  
Qy 1048 CLEIYVVVKGSLDQSLKD-----TLKFSI-----EKRPAWSGYVKS 1085  
Db 1587 -----YAWSSIDMSSTWFEIRIWHQHEKEBFTLKSNIIPQAGTSGSKGYPIFYIA 1641  
Qy 1086 LAHVAAETEGSSCSLLEYQMLV-----SARMLLIATTHADIMLTLTSVVRQQLFL 1138  
Db 1642 RRYKIGEING--DLIYHVLITLKPCHPFEFVVIDFTHTCSNDRFRTEFLQKPFVL 1697  
Qy 1139 DVL--DGTKALLVPAVSNGLSGMKCTLLILLRQKRELSGV-DEILGPL----- 1188  
Db 1698 PTVAIVENHAYIT--YNC-----NSWVREYTFHDIRILAPLKNRKL 1738  
Qy 1189 -----TEILEGVLODQ-----LMEKTKAVFSAFI-----TVLQMKEM 1223  
Db 1739 FLESFNKLTDPIDABQQLPGATILSLDEDLKVFSAALKSHKDTKVAIKVPTALQITSA 1798  
Qy 1224 KVSIDPQYSQVLAVCETLQEEVIALPQTGHSIALSGATDKDSMETDDCSRSRHRQR 1283  
Db 1799 EKTIKLAHSVILNIVVYASEIEVCLVDNDQFTLSTINES--GOLSFINDC----- 1848  
Qy 1284 DGVCTLGHLAK--ELCEVDEDDGSMQLQVTRRLPLP-----TLTTLTEVSL-RMKQNLH 1335

```

Db      1849 DNVVQAIHHRMRELSDPD-----STVHQIKRKPQVPGTLLMMLNMGSCDPNLR 1901
      1336 FFEATLHLLTLTARQOGATVAVAGCI-TQISCLPLSYQUSGTGTAQTPSARKSDA 1394
      1902 --TAAYNLCALATAPD--LKIEGQLLETQGLCIP-----SNNTFIKVSSEKL--A 1947
      1395 PMPGVYRLMSIMBQLKTLRYNPLPEALDPVGHQERTLOCUNAVTQSLACLLEAD 1454
      1948 TNEP---HTLTFLEESIOGF-----QRTTIELK-----LCLE--- 1978
      1455 HTVGFILQLSNFMKEMHFLPOLMRDIQVNLGYLCOACTSLHSRKLQHYLONKNGDGL 1514
      1979 -----YMTPW-----LKNLVKFCXSNDSKKLKVSQLDK-----L 2009
      1515 PSAVAGVORPPSAASAAPSSSKOPADPEASEQALH-TVYGL-----LKLSTLTAAL 1569
      2010 INLTIDQEMWPSVQAKTWGSIQIPELIDWVLDNLFHKSITYGSPQVEIMADTAVAL 2069
      1570 -----RHFTPDVCCIL-----LDQSL---DLAEY-NFLFALSPTTPTPDSE 1606
      2070 ASANVQVSKVITITICRMWDSCTNPQVYLEQHMMMDIALGRYLLMISFNCL---D 2126
      1607 VAPSGTLLATV-----NVALNMLGEIDKKEPILOAVGLSTQAEGRRT 1650
      2127 VATSVVYLFHTITPLVCSGSLSMRASTHGLVINIHTSLCTCTNP-----SFSEAGRV 2179
      1651 LK-SLMTFMENCYFLLI-----SOAMRYLDPADVPRDK-----QRMKQELSSLSLT- 1698
      2180 LRLSLDEFSLPK-FYLLFGISKVKAAYTAARSSCRHFTDKLGNERTVQPLPADRELS 2238
      1699 LSSL 1702
      2239 LPSL 2242

```

## RESULT 13

```

US-09-270-767-33392
; Sequence 33392, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33392
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-33392

```

```

Query Match      1.5%; Score 131.5; DB 4; Length 147;
Best Local Similarity 29.7%; Pred. No. 0.00015;
Matches 41; Conservative 25; Mismatches 61; Indels 11; Gaps 5;
      898 AAATRDALFRLQSKIEDMRKIMLEFLVAVETQGLIELFLN---LEVKDGSQSK 953
      1 ADQIRLTFMOKLPDELESDSIKIALIELVDACIAKQGVTEAFKXVAVLADKRSRSPFSK 60
      954 EF--SLGMSGLHAVLELIDSOODRYWCPEPLHRAAIAFLHALMQRDSAMLV--LRT 1009
      61 DCVFNIGE-SIVTYMRPFLDALQVDPLTIOQALPAKIMTIFHSWM--KHNLOMLVDDLVK 117
      1010 KPKFMENTSPPLFGTSLP 1027
      118 DKQFWKLCSPFSELOP 135

```

RESULT 14  
US-09-270-767-48609

```

; Sequence 48609, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48609
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-48609

```

```

Query Match      1.5%; Score 131.5; DB 4; Length 147;
Best Local Similarity 29.7%; Pred. No. 0.00015;
Matches 41; Conservative 25; Mismatches 61; Indels 11; Gaps 5;
      898 AAATRDALFRLQSKIEDMRKIMLEFLVAVETQGLIELFLN---LEVKDGSQSK 953
      1 ADQIRLTFMOKLPDELESDSIKIALIELVDACIAKQGVTEAFKXVAVLADKRSRSPFSK 60
      954 EF--SLGMSGLHAVLELIDSOODRYWCPEPLHRAAIAFLHALMQRDSAMLV--LRT 1009
      61 DCVFNIGE-SIVTYMRPFLDALQVDPLTIOQALPAKIMTIFHSWM--KHNLOMLVDDLVK 117
      1010 KPKFMENTSPPLFGTSLP 1027
      118 DKQFWKLCSPFSELOP 135

```

## RESULT 15

```

US-09-540-236-3580
; Sequence 3580, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3580
; LENGTH: 1383
; TYPE: PRT
; ORGANISM: M.catarhalis
US-09-540-236-3580

```

```

Query Match      1.4%; Score 130.5; DB 4; Length 1383;
Best Local Similarity 18.9%; Pred. No. 0.013;
Matches 278; Conservative 200; Mismatches 513; Indels 483; Gaps 68;
      81 RISKFL-GLDEGSVOLLCYLOEDYGRGTSVYKVLQDEROS--QALLIKTADYYEER 137
      119 RIFGFKTSLNEQIIXI-----MQDDHRLPYDFIMTYDTQNGSIPHRILWRICE--EVS 171
      138 TCTLRGV-----LHLLTY-----FQDERHRYREYVADCVKLEKELVSKY---- 177
      172 SVFVRYLTHRPFWLHAWTHDPLPESIDEMITDKORFANDYSDS-DPTKMLIQYHDL 230
      178 -----ROOFELXY-----KTEAPTW-----ETHGNLMTROVSRW 207
      231 KILRYLMRQLFSVAVYRFTLEVRFAVALDGAGSVAAEAYKSLPNHLYLPTIOQIPRL 290
      208 FVQCLAREQSMLEIITLYAYAFEMASDU-----LVLTKNF 243
      291 ELQFLKRLSVHNLVHNLHFNPSKMEFADIVDKNWLATORIINPNVYLLKDYGHVLSRLG 350
      244 KEQFGSROTHNLVDETMDPFVDRIGYSALILVSGMDIESLHKCALDDR--ELHQRA 301

```



```
Db 351 KE-----SRFETAMLADMSGDY-----YEEK-----EGV-----NRRFETRT 386
Qy 302 QDGLICQMDCLMTFGDI PHHAPVLLAMALRHRLNEBESSVVRKIGTALIQUNVQY 361
Db 387 KDWOMODDFV-----SPAQESKSLIN-----OLKSDIMLBDEGATYKQMOOE 431
Qy 362 LTRLLQSLASGNDCTTSTACMCVYGL-----LSFVLTSLHTLGNODDIIDTACEVAD 417
Db 432 LVSAFKNSAFWEQAKTKTSQMLPLGLDKNDOSLRLPBLSTHACHSLKQLEEVAMITAR 491
Qy 418 PSLPELFWTEPTSGIILDSVCGMPHLL-----SPLLQL-----LR 456
Db 492 -----YLNENPN-----DOSCKRLSDIVVYLRPEAQAEDRLIFNDVGMGLKLP 538
Qy 457 ALVSGKSTAKKYVSFLDKMSFYNELYKPKPHDVISHEDGTLRRQTPGLPLPGLGQTLR 516
Db 539 AKITG-TASREVDSILMAAISGFYRLGAK-NARFPRDYVEMWTMPLYESLG-----LD 591
Qy 517 IPQGVGVMLD-----DRAVLR-----WEYSVSWTLPTCEIEMLHVSTADV 562
Db 592 FESVNRACQLDAGFRGFDAAHLAMTLNADLDYRS-----PSYALDRIV-----AGF 642
Qy 563 IQHCQRVPIIDLVHKVISTDLSDCLLPITSRIYMLQRLTVVISPPVDV-----614
Db 643 LTPCDSFSP-SLMHFPKMRDESFOEASLP-----LEGVTLADQIVLEALCLIHAGL 693
Qy 615 -----IASCVNCLTVLAARNPAKWTDLRHTGFLPVAHAPVSSLSQMSABGNANG 665
Db 694 REFDEYORLAPVLDMLHRIEHRVIRRYFKYKRETSQMHVIFKWNAGIASIRAN-----H 749
Qy 666 GYGNILMNSQEPQGEYVTIAF-LRLITLVKGQSGTOSQGLVPC-----VMFLXK 716
Db 750 HYHHTKNSGTQNTQIETLEFEVLESMASSQVSAKPACTITFAPRGALRVNPFGLT 809
Qy 717 EMLPSYHKRMYNSHGVREQIGCLLELIHAILMNCHEITLHSSHTSP--QFLC---ICS 771
Db 810 LMLD-----NMLSAFPRQDQARLDLMKAGLKRRGRDFLEDDDGAVFLDAILICGRDCA 863
Qy 772 LAYTAGCTVINIMIGVDTIDMWMAAOPRSDGAGGQGLITVYKLAASVTNNVIRL 831
Db 864 IFYTNV-----ADDTQOLRPASVSB-----LIEFFK-----890
Qy 832 KPSNVVSLPBOALSQGHAGNNILAVLAKYIYHGNDRALPRLATQILKRLATVAPMSVY 891
Db 891 ---SNV-----DWAVDLQADHTNQ-----ADQIAPKALTIAPKVLQ 923
Qy 892 ACLGNDAAAIRDAPLRLQSKIEDMRI-KVMILEFLTVAVETOPGLIELFILNEVKGSD 950
Db 924 QLLIYEHNA--NAPDASVFEYETKEVCEDEDLADFLQKI-----961
Qy 951 GSKPESLGMSCLAHVALELISQODRWCPPLHRAAIAFLHAIWDRDSAMLVLRK 1010
Db 962 -----LHA-----KOSQTYLPP--SPMNVKRLHLDQFPNOSPISL-PD 999
Qy 1011 PKFENLTPFLFTLSPPS-ETSEPSILETCALIMKIKICLEYVYVVKSDQSLKDTLK- 1068
Db 1000 ATTQMLABALQONLQSGQLEEVLTIGIOLPMSMPLNOIYHKLKPARAYLKHKLPL 1059
Qy 1069 KFSIEKRPAYMSGVYKSLAHHVAETGSSCTSLLEYQMLVSAWRMLIIATTHA----- 1122
Db 1060 LYDABDGLF-----EPLSLDGLGYOL-----KETLIDMAAQGVFDGMP 1099
Qy 1123 -----DIMHLTDS-----VPRQLPLVDLDTGKALLVPAVNCIRLQSMKCT 1165
Db 1100 MDEPDYHKLITDIYNNDDKASEFAHIVR-IYYD-----NLLPAGAN----- 1139
Qy 1166 LLLILRLQWKRELSVEIILGPLTEILGVLOADQOLMEKTKAKYFSAFI--TVLQMKEM 1223
Db 1140 -----RLTSLDNLVGLAKLTDPAONLTDLOVKKQOLOYSMISPTKQYIAY 1188
Qy 1224 KVSQDIP-----QYSQVLVANCET-----LOEVIYALFDQTRSLALGS 1261
Db 1189 ELNHIPLDIYVAKLPISCDQWLRVLPNTVRAGHFLKFLYLHHLAMQVSRQTDQDIDIONLGS 1248
```

```
Qy 1262 -----ATEDXDMETDDCS-----RSNRD-----QRDGVCLGHLAKEICEVEDGDS 1306
Db 1249 SLMOFGQSLNDKDPHLDVSLKLPYHKTAKRQLOENFVWGL-LAKQ-----1297
Qy 1307 WLQVTRRLPILPILLTLE-VSLRMKONIAFTEA 1339
Db 1298 -----MPVLPLMTSLNIVALTAKNELFEGES 1324

RESULT 16
US-09-198-452A-113
; Sequence 113, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffls, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 113
; LENGTH: 1826
; TYPE: PR
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...1826
; OTHER INFORMATION: Xaa=unknown or other
US-09-198-452A-113

Query Match 1.4%; Score 129.5; DB 4; Length 1826;
Best Local Similarity 18.8%; Pred. No. 0.028;
Matches 360; Conservative 233; Mismatches 627; Indels 695; Gaps 93;

Qy 223 FLVAYFPMASDILLV--LTQPKKEQGRSGRQTRHVLNDEIMDPFVDRIGVFSALLIVEG 280
Db 92 FSHYSHATVSGSTELFSLHALLFTLEG---QADPKTEVLDLYSKERKVLSTIMELSEG 147
Qy 281 MDISLHKCALDDBRELHQFADGLICQMDCLMTFGDI PHHAPVLLAMALRHRLTNP 340
Db 148 VQISLIAPLAKDLAALHEVAQOQ-----FTKVCNGTINHITYFSLSGI--PE 194
Qy 341 ETSSVVRKIGTAAI--QANVFOYLTRLQSLASGNDCTTSTACMCVYGLSPVLTSL 398
Db 195 DCS--VDIVDTLLKSENNIARLKVSLPTALEFGHCS-----VLSDREL 238
Qy 399 HTLGNQDIIITFACEVADPSLPFLFWCTEFTSGIILDSVC-----GMFPHLSPLL 452
Db 239 MTFSTKQDID---VYTPPLTQQLF--SPHA-----LESRCSLCQSGGIFISIDNPL 286
Qy 453 QLRBALVSGKSTAKKYVSFLDKMS--FYNELYKPKPHDVISHEDGTLRRQTPKL--LYP 508
Db 287 -----IDENLSIKENCCSFAINCSSYLHTIYQALA-DALFNLLETPKXDSLPEIQNTF- 339
Qy 509 LGQGTNLRIPQGVGVN-----526
Db 340 LRGNKNTLVLPRLPDQTLGKNTLYKVRGVLANDIGDKVRYTTPRSRYLSKMSAHSCSL 399
Qy 527 -----LDDRAVLVNE-----YSVSWTLPTCEIEMLHVSTADVQHC-QRYKRP 571
Db 400 CKGTGLGIDYASVAVTEGKTTEFOQMSLNHNHVFPSKVS--PBLSTIQELIQLKQRLSP 457
Qy 572 IIDLHVHKVISTDLSDAC-----LPTSRIVMLQRLTVVISPP-----VD 613
Db 458 LIDGLGLVTPNRALATISGGEORTRIAKLHGLGELPEITTYILDEPSIGLHPDQTEKIG 517
Qy 614 VIASCVNCLTVLAARNPAKWTDLRHTGFLPVAHAPVSSLSQMSABGNANGYG----- 668
Db 518 VI-----KKLRDQGNITVILVEHEERMISLADRIIDIGPAGIGGGEVLF 561
```



QY 669 -----NLNNSPQGEY-----GVTAFLRL 690  
 Db 562 NKRPEDFLNNSSLTKYLRQELTIPPSREAPISWLLTATHTNLKLSIRPLAL 621  
 QY 691 ITTLVKGQIGSTQS-----QGLVPCWFLKEMLP--YHKRWYNSHGVEQIGLILELI 744  
 Db 622 IG--YVGVSGSGKSLINNTLVPAISFLKQENPKMLHPEW-----GC-IGNLI 667  
 QY 745 HAILNCHETDLHSHSTPSPLOCLCSLAYTEAGQVINIMIGVDTTDMMAOPRS-- 802  
 Db 668 HT-----TRDLPGSQRSI-----PLTYIKA-----PDDIRELPASQPRSLR 704  
 QY 803 -----DGA--EGQCGQLIKTVKLAFTVNNVIRLKPNSVSPLE-----Q 843  
 Db 705 QGLTKAHSPFNQPGACICQGLGIMTI-----SDDTPIPCSECQ 745  
 QY 844 ALSQGA-----HNNLIATVAKYIY-----HKHDPALPRLAIQMLRLATV 885  
 Db 746 GKRYHSEVLEILYEKNINADIDMTAYEAKFPFISHPKHEKIALCSLRID----- 797  
 QY 886 APMYVYACIGNDAAIRDAFLRLQSKIE---DMRKMILFEITVAVERO--GLIEL 939  
 Db 798 -----YLPGRPLSTLSCGEIQLKLAHELLFASPRQTLVYLDDEPTGLHTHDIOALIEV 852  
 QY 940 FLNLEVKDQSDGSKESPISLGMWSCLHAYVELIDSQODRYWCPPLHRAAIA-FLHALMOD 998  
 Db 853 ILSTLY-----LG-----HTVLYIHN-----MHVVKCDVYELQPE 885  
 QY 999 RRD-SAMLVLRTPKFKFMENTLSPFLGTSPSEIS-----EPSLLETQALIMKTIC 1048  
 Db 886 GGDGLGAYLLASCTPPDLIQLNTPAKALAPYEGSLDIPVYKSEPPSPKSCDILIK--- 942  
 QY 1049 LEIYVYVKGSLDOSL-KDTL-----KFSIEKRFAYWSG-----YKSLAV 1088  
 Db 943 -DAYNNMLKHIDMLAPRSILAIAGPGASGHSIVFDLYASGAINAYALEPPYIRQGL 1001  
 QY 1089 -----HYAETEGS-----CTSL-LEYQMLVSAM-----RMILLIATTHADIMHL 1127  
 Db 1002 KETPLPSGVEYKGLSPVISVKKSSNSRYHTIASALQSLNGLEKLFALILEPPSPLEE 1061  
 QY 1128 TDSVVRQLFLDVL-----DGTAKLLVPASVNCRLRSGKCTLLILRLQWRE--- 1177  
 Db 1062 KLSKTPPTITIDSLKSKYKDYVITTSPIP-----LGSD-----LELFQEKKEGEGFIK 1110  
 QY 1178 -----IGSVDEILGLPTEILEGVLOADQUMKTKAVFSAFTTVLQMKMKVSDIPQY 1231  
 Db 1111 LYSEGNIYDLBERL-PLNLEPAIVIOHTKVSPPKSSLSAISVARSLSSEIWIYISQK 1169  
 QY 1232 SOLVIANVETIQEVIYALFDQTRSLAGSATEDK-----DSMETDCSRSRHRDQDV 1286  
 Db 1170 KQKLSYSYLGWKDKKGRLYPETHQLXSDHPEGRCLTCGGRGELIKLSIEHKEKTAHY 1229  
 QY 1287 CVL---GLHLAKELCEVDEDEDSWLOVTRRL-----PILPTLLTTLTLEVSLEMKONLH 1335  
 Db 1230 TPLPFFSLFPFK-----SYMKYQKLLKQENASQPL--KLITL-----KEFLN 1270  
 QY 1336 FTEATLH-----LILTARTQO-----GATVAVAGITOSI-----CLPPL 1371  
 Db 1271 FCRSSSEPPGNMALLMEQDLETESDPLIKPLALTSCPACKSGSLANDYANVVRINTSL 1330  
 QY 1372 SVYQ-----ISTGTQTPS-----AARKSL-----DASWPCVYR 1402  
 Db 1331 DIVEEDATLESFLNTIGTDTKRSIIODLMNRLFTISKVGSYITLQORODTSLDGENYR 1390  
 QY 1403 LSNMS-----LMEQLLKLRYNPLPEALDFGVHQBERTLOCLNAVTRVQSACL 1450  
 Db 1391 LHLAKKISIMLNTIVLPEEPFLSGHPODLPTIYQLL---KELVANNVYIAIDRSGSLI 1447  
 QY 1451 EEAADHY-----GFIQLQSNFM---KEWFHLPQLM----- 1478  
 Db 1448 PHADHAFILPGSGPGGFLMDSPTVEVCPSYDLHANVPQTEVCPKAPLSISKANHTGSD 1507  
 QY 1479 RDIOVNLGYLQOACTSLHSRKMQLQHTVLQNKNGGGLSVAQVQRPSPA--ASAPSSS 1536

Db 1508 RTLKVNLSI-----HHION-----LKYSAFLHALVALVAGVSGS 1540  
 QY 1537 KQPADTEASEQALHTVQV-----LTKLSKTLAALRHFPDVOQIILDQSLDAEY 1590  
 Db 1541 GKTSLLEGGFKQABELIAKTTTFPSDLVIDSHPIAS-----SQRSIDSTY 1587  
 QY 1591 NFLPALSFTPTFPDSEVAPSPFGTLLATVN--VALNMLGEL--DKKEPLTOAVGLSTQA 1645  
 Db 1588 -----FD--IAPSIKAFYASLTQAKALNISTMSTFTWKQCCSDCGGLGYQ- 1632  
 QY 1646 EGTRTLSLMTFTEWENFYL-----ISQARY----- 1673  
 Db 1633 -----WIDRAFYALBKRPCTCGFRIQPLAQEVLYEGKHFGELHTPIETVA 1680  
 QY 1674 LRDPVAPRDKQKQKQELSSSELSTL-----LSSIS-----RYPRGAPSSP 1714  
 Db 1681 LRFPFIKKIQKP-LKALLDGLGYLPICQKLSLSVSEKTKALTKAVLYQTPETP 1734

RESULT 17  
 US-09-538-092-1154  
 ; Sequence 1154, Application US/09538092  
 ; Patent No. 6753314  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Glac, Inc  
 ; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
 ; FILE REFERENCE: 15966-542  
 ; CURRENT APPLICATION NUMBER: US/09/538, 092  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: 60/127,352  
 ; PRIOR FILING DATE: 1999-04-01  
 ; PRIOR APPLICATION NUMBER: 60/178,965  
 ; PRIOR FILING DATE: 2000-02-01  
 ; NUMBER OF SEQ ID NOS: 1387  
 ; SOFTWARE: CuraPatSeqFormatter Version 0.9  
 ; SEQ ID NO 1154  
 ; LENGTH: 3210  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: (0)...(0)  
 ; OTHER INFORMATION: Polypeptide Accession Number P49454  
 ; US-09-538-092-1154

Query Match 1.4%; Score 129; DB 4; Length 3210;  
 Best Local Similarity 19.2%; Pred. No. 0.092;  
 Matches 361; Conservative 282; Mismatches 684; Indels 550; Gaps 99;  
 QY 18 RELMTLLGRSALRELSQLEALNKGRRLLEGLSTYKPPSPSAAKRVANKQVAPLKE 77  
 Db 680 RNHLNVLDSKS-----VEVETQK-----LAYMELQQAEPDQKHQKEI---EN 720  
 QY 78 LGLRISKFLGL--DEEQSVOLL-----QCY--LOEYRGTRDSVK-----TVLQDERO 121  
 Db 721 MCKKTSQQLGQVEDLEHKLQILSNELMDKRCYQDLHAYESIRDLKSKDASLYVNEH 780  
 QY 122 SQALLIKINDYYEERTCILRCVHLL---TYFQDRHRYVE-----YADCVDK 168  
 Db 781 ORSLTA-----FDQOPAHNSFANIGRGSPSPERSSECRLEADQSPKNSAIIQNRVDS 834  
 QY 169 LE-----KEVSKYRQOFEELYKTEAPRWETHNGVLMTERQVSRMFV-----QCLAR- 214  
 Db 835 LBSLSLQKQMSDIOKQCEBELVQING--EIEENIMKAEQNHQSFVAETSGRISRLQED 891  
 QY 215 -----QSMLEIIFLYAAPEMAPSDILVLT-TPKPKQO--FGSRQTNRLHVLDETMDPFVDR 268  
 Db 892 TSAHQVVAETL-----SALENKEKEQLQINDKVEYEQEAIQGLKXSNHLBESL-----XE 943  
 QY 269 IGFYSALLIVEGMDESLKALKLDRRELHQAQGLICQDMDCMLTFGDIPIHAPVLL 328

```

Db 944 LQLSETLSEKEMSIISL---NKREIELTQENGTLKEINA----- 984
Qy 329 AMALLRHLPBETSSVVRKIGTAIQLVNPOYLRLRLQSLASGNDCTTSTACWKG 388
Db 985 -----SLN-OEKNNLIQKSESFA-----NYDEREKESIS----- 1012
Qy 389 LSFVLSTLEHLTGNQOD-IIDTACEVL--ADPSLPFLWGT-EPSTGILIDSVCMF 444
Db 1013 -----ELSDQYQEKILILQRCBETGNAYEDLSQTKKAAQKNSKLECLINACTSLC 1064
Qy 445 PHLSPLLQLLRALVSGKSTAKKYSFLDKMSFYNELYKHKPHDVISHDGTLWRQTPK 504
Db 1065 ENRKKEELQLKEAF-----AKEHQEFLTKLAFABE-----RNQ--- 1097
Qy 505 LLYPLGGQTNLRIPGTYGQWV---LDDRATVVRREYSSWTLTCEILEMLHVVSTAD 561
Db 1098 -----NMLELETVOALRSEMTDNQNNKSKEAGLKOELITLKEONKQKEVND 1148
Qy 562 VIQHCQVRKPIIDLHVKVIISTDLSTADCLPITSRIYMLQRLTTVISPPVDIVASVNC 621
Db 1149 LIQNEQOLMKVKTGH-----EC-----QWLE---SEPI----- 1174
Qy 622 LTVLAARNPAKVWTDLRH-TGFLPFVAHPVSSLS-----QWISAGNNAGGYNMLMNS 674
Db 1175 -----RNSVKEBERERQCNPKQMDLEVKESLSDSYNAQLVQLEAM-----LNK 1220
Qy 675 EQPGEGVVTIAFLRLITLVKQGLSGTOSQGLVPCMFVKEMLPYHKRKYNSHGVRE 734
Db 1221 ELKQESKEKEKCELOHBLQTRIGDLETSLNLDQMOQOESIGLKD----- 1263
Qy 735 QIGCLILELHAILCHETDLHSHTPSLOFLCISLAYTEAGOTVININGIVDTIDM 794
Db 1264 ---CEI-DAEKYISGPHELTSQNDNAHLQ---CSL-----QTNMKLN-ELEKICE 1308
Qy 795 VMAAQPRSDGAE-GQGQGLIKTYKLAFSV---TNVYIRLKPSS---NVVSP--- 841
Db 1309 ILQEKYELVTELNDRSRECIATATRKMAEVCGLNBYKILINDSGILHGEIVEDIPGE 1368
Qy 842 ---EQALSHGAGNNLIVAKYIYHK---DPLPLRLAIQLKRLATVAPMSYACLG 895
Db 1369 FGEOPNEHPVPS---LAPLESNSYEHLTSDKQVOMHPALQKFLSTL---QSHKIL- 1421
Qy 896 NDAARIPAFRLRLQSKIEDMRKIMLEFLTVAVETQPGILEPLNTEVDKSGSGKEP 955
Db 1422 HDQHCQSSKXSELOTYVDSLKAENLVLS-----TNLRNPG-DLVKEM 1464
Qy 956 SLGMSCLHAYLELIDSOQODRYWCP---LHRAAIAFLHMLQDRRDSAMVLRTKP 1011
Db 1465 QLGIEBGL--VPSLSS---CVPSSSLSSLGSSFYRLLEQTDMSIL----- 1509
Qy 1012 KFWENLTSPPLFTGTLSPSETSEPSILETCALIMKIIIC--LEIYVYKGLSDQSLKOTLK 1069
Db 1510 -----SNLEGAVS-----ANQCS-VDEVFCSLQTYVDSLKAENLVSTNRN 1551
Qy 1070 FS-----IKRPAVMSGVYKSLAVHVAETEGSSCTSLLEYOMLVSMRMLIIATTHADIM 1125
Db 1552 FQGDVLEMQGLBGLVPSLSSCV--PDSSLSLGD---SSFYRLLEQTDMSILS 1606
Qy 1126 HLTDSVVRQLFLDVLDTGKALLLVASVNCILKLSMKCTLLILLROMKELSGVDEIL 1185
Db 1607 NLEGVVANQCSVD-----EYFCSSIQEENILT-----KETEPAP--- 1641
Qy 1186 GPLTEILEGVLOADOQLMEKTKAKYFSAPITVLOKEMKVSIDIPOYSQVLNVCTLOE 1245
Db 1642 AKGVEELESLCEYVQSLKLEKEMESOGI--MKNKIEIOELBOLSSERQELCLKQ 1697
Qy 1246 VITALFDQTRHSLALGATEDKDS-----METDQSRSRH--RDQRDV-----CVLGIH 1292
Db 1698 YLSENEQWQOKLT--SVTLEMSKLAAEKQTEQSLLEVARLOLOGIDSSRSLGLID 1755
Qy 1293 L-----AKELCEVEDDEDSDMLQVTRRLPIPLTLTLEVSLRMKNLH---FTBATLH 1343
Db 1756 TEDAIQGNESCDISKEHTS--ETTERTP-----KHVHQICDQKQAOQL 1798

```

```

Qy 1344 LITLAR--TQAGATAVAGITQSTICLPLLSYQJSTNGTQATQPSARSLDAPSWGYR 1402
Db 1799 NLDIEKITETGAVKPTGE-----C-----SGEOSPDTNYP-----PGEDK 1834
Qy 1403 LMSIMEQLKTLRLNPLPEAL---DFVG-----VA--QERTLOCLINAVTQSLACLEBA 1453
Db 1835 TQGS--SECSISELSSG--PNMLVPMDFLGNQEDIHNLQARKYEN--ENLRHLHVIDR 1889
Qy 1454 DHTVGFILQSNFKE-----WHFHLPOLMRDIOVNLGYLCOACTSLHSRMYLQHYLQNK 1509
Db 1890 DRKVESL---NEMKEILDSKHLQEVQIMTKI-----EACIEL-----EKIVELKKE 1934
Qy 1510 NGDGLP-----SAVQORVQRPASA---AAPSSQOPADTER---SQOALH 1552
Db 1935 NSDLSEKLEYPSCHQELQEVETSEGINSDLEMHADSRBEDGDNAVKNDWKERPL 1994
Qy 1553 TVQVGLKILSKTLAALRHFTPDVQCIILDOSLDLAEYNPLFALSFTTPTPDE-----VA 1608
Db 1995 DVENELSRIRSEK-ASIHHA-----LYLEADLEVVOYEKLCLEK-----DNEHKQYI 2042
Qy 1609 PSFGLTATVNAVLMNL-GELDKKKEPLTQAVGLSTQ-ABGTRILKSILMFTWENC FYLL 1666
Db 2043 VCLREELSVTSSERQLEGLDMSKKTALDQLEKKEKETOELSHQ----- 2091
Qy 1667 ISQMRVYRDBAVHPRDQKQKQELSSLSLSTLSSLSYFRGAPSSPATGVLPSPQGS 1726
Db 2092 -SECHLCIQVAEVEKTELLQTLTSDVSELKDKTHQEK-----LOSLEKDS 2140
Qy 1727 TSLKASPEQOPLIQL 1743
Db 2141 QALSTKCELENOIQL 2157

RESULT 18
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
;
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
US-08-353-700-1

```

Query Match	1.4%;	Score 129;	DB 1;	Length 3248;
Best Local Similarity	19.2%;	Pred. No. 0.094;		
Matches 361;	Conservative 282;	Mismatches 684;	Indels 550;	Gaps 99;

QY	18	RELMTILIGSALRELISOIEALINHMRRLLBGLSYKRPSPSSSAEKVANKDVASPKKE	77
Db	680	RNLHNVLDSDKS-----VEEVETOK-----LAVMELQGAEPSPDQKHQIEI-----EN	720
QY	78	IGLRISKPLG--BEBOGVOLL-----OCY--LOSDYGTBPSVK-----TYLQDBRK	121
Db	721	MCLTSGTQVEDIEHKLQJLSINEMDKRCYODLHAEYSRLRLKSKDASLTVDNDA	780
QY	122	SOALILKIADYVEERTCILRCVLLH---TYFODERAPVEVE-----YADCVDK	168
Db	781	QRSILLA-----PQOPAMHHSFANIIQEGOSMPSESECHLEADQPSKALLQNRDS	834
QY	169	LE-----KELVSKYKROQFEELKYTEAPWTETHGMLMTEROVSRYFV-----QCLRE	214
Db	835	LEFLESQKONSJDLQXCEELVOIKG--SIEENIMAEOMHOSFVAETSQRI SKIOED	891
QY	215	-----QSMLEIIFLYAYFEWAPSDLLVLT--KMEKQG--FGSRQNRHLVDETMDPFVDR	268
Db	892	TSAHQNVVAEVL-----SALENKEKEKELQJLNDKVELEQAEIDELKSNHLEBDSL--KE	943
QY	269	IGYRSALIVGMIIESLHKCALDRREHLQFADQGLIQQDMCMLITFGDIPIHNAVYL	328
Db	944	LOLISETLSEKKMSSIIISL--NKREIEBELTENGTLKEINA-----984	
QY	329	AMALLRHTLNEEFSSVVRKIGTAFIQLNVQYVLTLLQSLASAGNDCTTSAGMCVYGL	388
Db	985	-----SLN-QEKMNLQKSESFA-----NYIDEREKSI.S-----1012	
QY	389	LSFVLTSLBHTLGNOD-IIDTACVYL--ADPSLPELFWGT--EPTSGLAGIIDSVOGMF	444
Db	1013	-----ELSDQYKOEKLILQRCSEETGNAYEDLSQYKAAQKNSLTBCLTNECTSLC	1064
QY	445	PHLSPLLOLRALVSGSKTAKKYYSFLDKMSFYNELYKHKHPHUVISHEDGTLRRQTPK	504
Db	1065	ENKRNIEELQKEAR-----AKEHOEFLTYKIAFEAE-----RND--1097	
QY	505	LIVPLGQOTNLRIPQGTVGQVM--LDDRAVLVMEWESYSWILFTCEIEMLLHVSTAD	561
Db	1098	-----NLMLELFTVQQLAREEMTDNQNNSSSEAGLQKQELMTLKEBQNKMQKEVND	1148
QY	562	VIOHQCRVKRILDLVHKVISTDLSIADCLLPITSRIYMLQRLTTVISPPVDVIASCVC	621
Db	1149	LLOENBELMKVMYKTKH-----BC-----QNE--SEPI-----1174	
QY	622	LTVLAANPAKVMYTDLRH--TGFLPFWAHVPVSSLS-----OMISABGNAGVGNILMNS	674
Db	1175	-----RNSVKERSEBENQCNFKQMDLEVKELISLDSYNAQVLOJLDM-----LRNK	1220
QY	675	EOPQEGYGVITAFIRLITTVKQGLSTQSGVLVPCWFWLKEMLPYHKRYNSHGVR	734
Db	1221	ELKLQSEKEKECLOHQLQIRGDLFTSNIQDMQSOEISGLKD-----1263	
QY	735	QIGCLILIELHALINLCHEITDLSHSHPSLQFLCISGLATVTEGQTVINIMIGIVDTIM	794
Db	1264	-----CEI--DAEKKYISGPHELSTQNDVAHQ-----CSL-----QTTMANKLN--ELEKICE	1308
QY	795	WMAQPRSDGAE--GQGGQOLLIKTVKLAFSV--TNNVIRLKPSS-----NVVSPL---	841
Db	1309	IIOAEKVELTEVLEINDSRSECITATRKMAEEVGKLLNEVKTILNDDSGLLHGEIVEDI	1368
QY	842	-----BOALSQCHAGHNNLIVAKTYHKG-----DBALPRALIOJLKLKATVAPMSVYACLG	895

Db	1369	PEQCNQEHQPVSS----	LAPLDESNSEYHETLSDKEVQNHFAELOEKFLSL--QSEHNKL-	1421
Qy	896	NDAAAIRDAFLTRLOSKJEDMRKIMWILEFLTAVENTOPGLIELFMLEVXGDSGSEF	955	
Db	1422	HOHCQMSKXSELOQTYVDSLKAENLVLS-----	TYLRNFOG-DLVXEM 1464	
Qy	956	SLGMSXCHAVLEILDSQODRWCP-----	LHRAALAFHAIAMORBRDAMLVLTKP 1011	
Db	1465	QJLGEELG--VPSLSS-----	CVDPSSLSLSDGSFFRALLEOTGDMSSL----- 1509	
Qy	1012	KFWENTSPLEGLTSPSETPSEPILETICALIMKIC--LEIYUVKXGSLDOSLKTCLK	1069	
Db	1510	-----SNLGAVS-----	ANQCS-VDEFVCSLSQTYVDSLKAENLVSTLYRN 1551	
Qy	1070	FS---IKRPAVNSGVYKSLAIVHAETEGSSCTSLLEYOMLVAMRMLLIATTHADIM	1125	
Db	1552	FOGDLVKMOMGLEELGVLPSLSSCV--PDSSLSLSD-----	SSFPYRALLEOTGDMSSL 1606	
Qy	1126	HLTDSVVRLOFLFVLDGTAKLLLVPSVYNCLRLSGSKCTLLILLQMKRELGSVDEIL	1185	
Db	1607	NLEGVVSNQCSVD-----	EVFSSLSQENLT-----RKETPSAP-- 1641	
Qy	1166	GPLETELEGVLADQOLMEKTKAKVFSAFITVLQMKEMKYSIDIPQVQLVLMNCELOE	1245	
Db	1642	AKGVSELESLCEVVRQSLSEKLEEKMEQSGI---	MKNKEIQELEQLLSSERQELDCLRKQ 1697	
Qy	1246	VIALFPDQNRHSLAGSATDEKDS-----	METDDCSNRH--RDQDGV-----CVLGLH 1292	
Db	1698	YLSNEBQOQKLT--SVYLEMESFLAAEKQOTBOLSELEVARLOGLDLSRSKSLGID	1755	
Qy	1293	L-----AKELCEVEDGDSWLCQTRRLFLPLTLTLEVSLMKOVLH--	FTEATLH 1343	
Db	1756	TEDAIQGRNECDLSKHTS--ETTETP-----	KADVHOICDKDAQDL 1798	
Qy	1344	LTLTAR--TOOGATVAVAGITQISICLPLSVYOLSTNGTQTPSARXKSLDABSPGVVR	1402	
Db	1799	NLDIEKITEGAVKPTGB-----	C-----SGEQSPDYNVR-----PEDEK 1834	
Qy	1403	LSNMLMOQLKTLRYNLPREAL--DPVG-----	VH--QETTLQCLANVRVYOSLACLEEA 1453	
Db	1835	TOGS--SECSISELSFG--PNALVPMPLGQOEDIHNLQLEVKETSN--	ENILRLHVLIEDR 1889	
Qy	1454	DHTVGFILQISNFMKE-----	MHFLPOLYMDIQVNLQOACSTSLHSKXKLOHYLQNK 1509	
Db	1890	DRKYESIL---NEKKEIDSKLHIOEVOLMKI-----	EACIEL-----EKIVGEIKKE 1934	
Qy	1510	NGDGLP-----	SAVAQRVORPSPAS-----AAPSSSKOPADTEA-----SEOOALH 1552	
Db	1935	NSDISEKLEFVSCDHQELQVETSEBANDLEMHADKSRREDIGDNVAVNWSKWERFL	1994	
Qy	1553	TYOQGLKILSKTLAARHTPVQCQLLDQSLDLAEYANFLFALSTFTTPFDE---	VA 1608	
Db	1995	DVEEELSRISSEK--ASIEHBA-----	LYLEADLEVQVTEKLCLEK-----DVENKOQVI 2042	
Qy	1609	PSFETTLATVAVNLMLM--GELDKKEBFLTOAVGLSTO--	AEGETFLKSLMFTWENCGYLL 1666	
Db	2043	VCLSEELSVYTSERNQJRGSLDIMSCKKTYLDLDSKEMKEKTOBELSHQ-----	2091	
Qy	1667	ISQAMRYLRDPVAVPRDKQMKQKLSSELSLSSRYFRRGAPASSPATGVLPSPQGS	1726	
Db	2092	SEGLHCIOVAEABVKEKTEYLQTLSSDVSEILKDKTHLOEK-----	LQSLKXS 2140	
Qy	1727	TSLSKASPEQEPLOU 1743		
Db	2141	QALSTLTCLELNOIAOL 2157		

RESULT 19  
PCT-US95-16216-1  
; Sequence 1, Application PC/TUS9516216  
; GENERAL INFORMATION:  
;     APPLICANT: Yen, Timothy J.  
;     APPLICANT: Ratner, Jerome B.



QY 1454 DHTVGEIILQSNFMKE-----MHFHLPOLMRDIOVNLGYLCOACTSLHSRKMLOHYLQNK 1509  
 DB 1890 DKRVESL---NEMELDSKHLQSVQMTK-----EKICEL---EKIVELKKE 1934  
 QY 1510 NCGDLP-----SAVQVQRPSPSA-----AAPSSKOPADTEA-----SEQQALH 1552  
 DB 1935 NSDLSEKLEVFSCDQELLQRYVETSEGLNSDLEMHADKSSREDIDQNVAKVDSWKERFL 1994  
 QY 1553 TVOYGLKLTSLKTLAALHFTPDVCOIILLDOSLDLAENFLFALSTFTPTDSE-----VA 1608  
 DB 1995 DVNELSRIRSEK-ASIEHEA-----LYLEADLEVQTEKCLEK-----DNEKQKQVI 2042  
 QY 1609 PSFGTLLATVNVALNML-GELDKKKEPLTQAVGLSTQ-AEGTRTKSLIMFTMENCYLL 1666  
 DB 2043 VLEELSELVYTERNQRLGEIDTMSKKTALDQLEKMKETQELSHQ-----2091  
 QY 1667 ISQAMRYLDPVHPDQKQKQELSELSTLSSLSRYFRGADSSPATGVLPSPQGS 1726  
 DB 2092 -SECHHCIOVAEAYEKETELLQTLSSDVSELKDKTHQEK-----LQSLERDS 2140  
 QY 1727 TSLSKASPESEPLQL 1743  
 DB 2141 QALSLTKCELENOIQL 2157

RESULT 20  
 PCT-US94-00198-3  
 ; Sequence 3. Application PC/TUS9400198  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schering Corp.  
 ; TITLE OF INVENTION: RAS Associated GAP Proteins  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Schering Corp.  
 ; STREET: 1 Giraldo Farms  
 ; CITY: Madison  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 94304-1104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: Macintosh  
 ; OPERATING SYSTEM: 6.0.8  
 ; SOFTWARE: Microsoft Word 5.1a  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US94/00198  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/004,824  
 ; FILING DATE: 15-JAN-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lunn, Paul G.  
 ; REGISTRATION NUMBER: 32,743  
 ; REFERENCE/DOCKET NUMBER: DX0352 PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (201)822-7255  
 ; TELEFAX: (201)822-7039  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2938 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Saccharomyces cerevisiae  
 ; PCT-US94-00198-3

Query Match 1.4%, Score 128.5; DB 5; Length 2938;  
 Best Local Similarity 19.2%; Pred. No. 0.088;  
 Matches 183; Conservative 149; Mismatches 351; Indels 271; Gaps 41;

QY 971 DSQOQDRYWCSPPLHRAI-AFLHALMQDRDSMLVLRKPKKMENTSPLETSPPS 1029  
 DB 807 ERSQSPSSPPQOQSDLPSPUSLVSSAGFSNNHSITATPTLLKNIKSP-----KPN 860  
 QY 1030 ET-----SEPSILETCALIMKIICLEIYYVVKSLDQSLKDTLKKKSIEKRPAW- 1079  
 DB 861 KTKIADKQKQKQSYR-----VILSDNEARKIMNITSIFRGMNWF 905  
 QY 1080 -----SGYKSLAVHAETEGSSCTSLLEYQMLVSAWMLLIATTHADIMH 1126  
 DB 906 IRPDANTEPKFTFDILIKPLFVSLIDS-----NQRLQVAPAFIEIPLSYIAFED 956  
 QY 1127 LITSVAROLFVDLTOSTKLLIIPASVNCRLGSMKCTILLILRQMKELSYDEITG 1186  
 DB 957 IDNDLDFRVLNDHYLCTTAVVTLFASLFDPLKENARREMLDIIVKFOR-----VASYLS 1012  
 QY 1187 PLTE---ILGVLQADQQLMEKTKAKVFA-----FTVLOMK 1221  
 DB 1013 NLAEMKHLVQAIIITTELTLPLVGVAGSGIFISYCSRGNTPLIKISCEPLRSIRFY 1072  
 QY 1222 EMKYSDIPOYSQVLVAVCEYLQEE-----VIALPDQTRHSLAL--GSATEDKDSMET- 1271  
 DB 1073 QKYVGALDOYSIYVIFIDAMADNFTASGSVALQRLRNIIITLYIKGSDSILDSMDVI 1132  
 QY 1272 -----DDGSRHRDQ-----RDGVCVLGLHLAKELCEVEDGDSMLQVTR 1312  
 DB 1133 YKMFYFSCSKSYTOEELVDFRSLAGILASMSGILSDMOELSKSKAPDNDEGS-LSFES 1191  
 QY 1313 RLPLPFTLLTLEVSILMKONLHFTATYHL-----LTLTARTQOGATAVAGAGITOSICL 1368  
 DB 1192 RNPAV-EVHKSLKLELTKKGNFPISSQOCWLNPNLITLHNSD-----LSTIEL 1240  
 QY 1369 PLISVYQLSTN-----GTAQTPSASRSLDAPSPGVYRLSMSIMQLKTLR----- 1416  
 DB 1241 HPLSFNLFNNLGLKIDELMSIDLSKSHDS-----SFVLRLQIILIIIFTLIKRSD 1291  
 QY 1417 -----YNFLPEALDFVGHQERTLOCLNNAVTVQSLACEADHTGFTILOSNFMKEW- 1470  
 DB 1292 DEKIMLLFSTDLDVAD-----KLIEIYERKS-IKSKRYYG-IIQMSKMPRAE 1339  
 QY 1471 -----MHFLPOLMR-----DIOVNLGYLCOACTSLHSRKMLOHYLQNRKNGDG 1513  
 DB 1340 HSEKNLGISNHFPLKKNMKLVLIGWFKLSINKQYDPEN-----SPLREMDLQKDEDF 1394  
 QY 1514 L-----PSAAYQVQRPSPSAASAAPSSKOPADTEASEQALHTYQY-----LTK 1560  
 DB 1395 LYIDTIESAKALAYLTHNP--LEIPSSSKE--DMNRSS-----TVSFGNHTILIK 1444  
 QY 1561 ILSKTL-----AALRHFTPDVCOIILLDOSLDLAENFLFALSTFTPDSEVAPSGTLL 1614  
 DB 1445 GLEKSAIDNQFPVSLHKI SILNENVIATLNLNSNANVAVSLKFTLP---MGSPKQDR 1501  
 QY 1615 LATVNVALNML-----GELDKK-----KEPLTQAVGLSTQAEGRITLKS- 1653  
 DB 1502 IAFRLVFIIVTVYVNPVPEKHEMDKMLAIDFLKYLIIKNPIIAFFGLSLACSPADVLDVAG 1561  
 QY 1654 --LIMFTMENCFLILISQAMRYLDPVHPDQKQKQELSELSTLSSLSRYFRGADSSPATGVLPSPQGS 1726  
 DB 1562 GFIMAFDTRASHITVTELL-----KQETKRAARSDDIIRNSCATRALSLVT 1609  
 QY 1707 RRGAPSSPATGVLPSPQGS-----KSTSLSKASP--ESQEPILQLVQAVFRHMQR 1753  
 DB 1610 RSRGNKYLIKTLRPLVQLGIVDKESFEIDKMKPGSENSEKMDL---FEKVTMR 1660

RESULT 21  
 US-09-538-092-1136  
 ; Sequence 1136, Application US/09538092  
 ; Patent No. 6753314  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Glot, Loic  
 ; APPLICANT: Manelfield, Traci A.

TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
FILE REFERENCE: 15966-542  
CURRENT APPLICATION NUMBER: US/09/538,092  
CURRENT FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 60/127,352  
PRIOR FILING DATE: 1999-04-01  
PRIOR APPLICATION NUMBER: 60/178,965  
PRIOR FILING DATE: 2000-02-01  
NUMBER OF SEQ ID NOS: 1387  
SOFTWARE: CurateSeqFormatter version 0.9  
SEQ ID NO 1136  
LENGTH: 3433  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (0)...(0)  
OTHER INFORMATION: Polypeptide Accession Number P46939  
US-09-538-092-1136

Query Match 1.4%; Score 127.5; DB 4; Length 3433;  
Best Local Similarity 18.6%; Pred. No. 0.15;  
Matches 389; Conservative 256; Mismatches 695; Indels 749; Gaps 95;

QY 8 TSVLSFCSRSREIMTILGRSALRELSQLAE-----LNKMRRLLEGLSYKKPSPSSA 62  
DB 364 SSVGVSLQAGNQLT-----QGTLSDEEFETOEQMTLNARNEAL-----RVESM 409  
QY 63 EKVYANKDVASPLKELG-RISKPLGLDEEOSVQLQCYLOEDYKGRDSVKTVLQDERQ 121  
DB 410 DROSLHVLWMLQKQQLQJLSAMLTLTEERIQKMETPLDDVY-----SLQKLEHKS 465  
QY 122 SQALILKADYYEERTILRCVHLILTYFODERHPYVEYADCDKLEKELVSKYRQF 181  
DB 466 LQSDI-----EAEQVKNVSLTHMVIVIDENSG-----ESATAI--LEDQI-----QKL 506  
QY 182 BELYKTAPEWETHNLMTEREQVSRMFVQCLREOSMLEIIFLYAVYEMAPSDLLVLT 241  
DB 507 GERW-TAVCRN-----TE--ERW-----NRLOEINILMOELLE--EQCLIKAWL 545  
QY 242 MFEQSGFSRQTNRLVDETDPFVDRIQYSALILVEGMDIESLHKCALDRRELHQA 301  
DB 546 TEKEBALNKVQTSNFKQKELSVSRLLA-----ILKEDMEK-----RQTLQQLS 591  
QY 302 QDGLICQMDCLMTFGDIPIHAPVLLAMALRLHNLNEETSSVVRKIGTRAIQLVNF 361  
DB 592 EIG--QVVGQL-----:-----DNSKASKKINSSEELT--QR 620  
QY 362 LTRLQLSLASGNDCTSTACWCYGLSPVLTSLHLTLG-----NOODIIDT-----A 411  
DB 621 WDSLQRLDESSNOYQAVNA-----KLGMSQIPQKDLLETAVRBEQA 662  
QY 412 CEVLADPRLPELFWGTEPTSGILILD-SVCGMPRLHLSPLLO-LRLALVSGKSTAKKY 469  
DB 663 ITKSKSKQLRP-----PPPKKQRIHVDIEAKKKFAISABELLNLKKTAKIQTLEIKY 718  
QY 470 SFLDMKSYNNELYKRPKPHDVISHEDETLMRQTPKLLYPLGGQNLRIIPQ-----GTVCQV 525  
DB 719 M--KMQDTSEM-----KKKLKALEKEQERIPRADELNQTQOI 754  
QY 526 MLDDRAYLVNREYSSWTLPTCEIEMLLHVST--ADVIOHCQVKRIIIL----- 575  
DB 755 LVE-----QMGKEGLPTEIKNVLEKVSSEKKNVSOHLEDEKRIQLOEDINAYF 804  
QY 576 -----VHKVIST-----DLSIADCLPITSRIYMLQR-LTTVVS--PPYDVI-ASGVN 620  
DB 805 KQDLBLEKVIKTEKEMVGHSTISSESSQSLPELXQSCQRELINLGLPKIEMARASQSA 864  
QY 621 CLTVLAAPNPAKVMTDLRHGTFLP-----VAHPVSSLQMTISAEKNNAGGYGNL----- 670  
DB 865 LMSQSA-----PDPVORGFDPSFLAGRYQAVQVBEIROOHLLENELKGGPQHALETLLX 917  
QY 671 -----LMNSEQPGGEVGTIAFLRLITTLVKGQLSGTQSGVLPCVMFVLKEMLPYHKMR 726

DB 918 TLKQVNLDSNKQAVS-----LVNLNDLAVKEXALQEKTLDE-----ILENQKALMKLA 968  
QY 727 YNSHGREQIGCLILEIHA-----ILNICHETDLSSHTPSLOFLCISLAYTE 776  
DB 969 EETKALEKKNVHPDEKLYKQEFDDVQKWNKLAKVLKSKDLHLEIAL-----TLRAFE 1022  
QY 777 AGQTVINIMIGV-DTIDMVAAPRSDGAEQ-QGQGLIKITVKLAFSVTNNV----- 828  
DB 1023 ADSTVIEKMDGVDFLMKQQAAGDDAGLQRLQDQSAFVNEIETLESSLKNKKEIETN 1082  
QY 829 IRLKPSNVVSPLEQALSOHGAHGNL-----IAVLAKYIYHNKDPALPRLAIQLIKRIATV 885  
DB 1083 LRSGPVAGIKITWQTRIGDYQTOLEKSKELATQKSLSSQEK-----AANLKQDLAEM 1137  
QY 886 APMSVYACLGNDAAAIIDAFITR-----LQSKIEDM-----RIKWMILE 924  
DB 1138 QEWMTQA-----BEEYLERDPEYKSPPELSSAVEEMKRAKEDVLQKEVRVKILKDN 1188  
QY 925 FLTVAVETQGLIELFLNLBYK-----DGSQSKFSLGMSCLHAVELIDSQQD 976  
DB 1189 IKLIAAVPSGGQELTSELNVLENYQLCNIRGKCHTLEEVWSCWIELHLHYLD----- 1243  
QY 977 RYWCPLIHRAAIAFLHALMODRDSAMLVLRTPKFEWMLTSPFGTLSPPSETSEPSI 1036  
DB 1244 -----LETTWNLTEBRMKSTEVLPKTDVANEAL-----ESLESVL 1280  
QY 1037 LETCALIMKTIICLEYVYVVGSLDOSLKDITLKRESIERFAYMSGYKSLAVHAETGS 1096  
DB 1281 RHPADNRQIHELQOTLIDGDIIDDIISEKLEAFN--SRYEDLS-----HIAESKQI 1330  
QY 1097 SCTSLLEQWLVSMRMLLIITATHADI-----WHLTDSV----- 1131  
DB 1331 SLEKQL--QVLRKTDQMLQVLSGLBLQKQTLTYLTDRIADFVPOBAQIOALEISANE 1388  
QY 1132 -----VR-----OLP-----LD-- 1139  
DB 1389 LTBELRRNRSQPLTSPBSTAGSQMDVLOKRLAEVSTKFPQKPAFQBRMDCK 1448  
QY 1140 -VLDGTALLIV--PASVNCRLGSMKCTLLILRLQMKRELGSV----- 1181  
DB 1449 RVLDGVAAELHVLVDKVDVDPVIOTHLDKCKIKYLTSEVGLBEVETVITGRHIVQKQOT 1508  
QY 1182 -----DEILGPI-----TEILBGV--LQAOQOLMEKTKAV-----FSAPITVQ 1219  
DB 1509 DNPXKMBQLTSLKVLVNDLGAQVTEGKQDERASQRLARRKKKAASLSEWLSFETELV 1568  
QY 1220 MK--EMKSDIPQYSQVLNVNCEILOEVALPDPQTRHSIALSATEDKDS-METDQC 1274  
DB 1569 QKTSBGLDGLDTEISMANKVVLKDLERKADNTITIBSSAALQNLLEGSPILBERLCV 1628  
QY 1275 -----SNSR-----HRDQR--DGVCV--LGLHLAKELCE----- 1299  
DB 1629 LNAQMSRVRTWTEDEMCNTLMNHQQLEIFPDGNVAHISTWLYQVADLDELKKEPTSQOE 1688  
QY 1300 -----VDEDEGSKVQV-----TRRL----- 1314  
DB 1689 IVKRLVSELDANLQVENVRDQALLMNAQSSSRREIPEKRLAELNRFKVSQHIKSAK 1748  
QY 1315 -----PIPLTILTL-----EVSIR 1329  
DB 1749 LLIAQEPYQCLVATTEFFETGVPPSDLEKLENDIENMLKPVKHLSESDDEKMDDESSAQ 1806  
QY 1330 MKQNLHFTFATLH-----LLITLART-----OQATVAVAGITQSTIC 1367  
DB 1809 IEEVLRGEBEMLHQPMEDEKNKEKIRLQILLHTRYNKIKAIPIQQRKMGQLASIRSSL 1867  
QY 1368 LPLISVQSLSTNGAQRPSARSKSLDAPSWGAVRRLMSLMEQLKTLRYNLFPEALDFV 1427  
DB 1868 --LPTDYIVETINKILLCMDDVELSLNVPBLATAIYEDPSFOEDSLKNIK-----DQLDKL 1920  
QY 1428 G-----VHOER-----TLQCLNA-----VRTVOS--LACLEBADHTVGF 1459



Db 1921 GEOIAVHEKOPVILASGPEAIQIRDTLTQLNAKDRINRMSDRKSCPDRA----- 1974

QY 1460 ILQLSFMKEM-HEH-----LPQLMDIQVNLGYLQACTSL-IHSRKLQHYLO----- 1507

Db 1975 -----MEMQFHDNDLNTQWITEABELLVDTCAFGSLDLKRAKHQOELVVGIS 2027

QY 1508 -----NKNDGCLPSAVACRORPPSA-----ASAPSSSQPADTEA 1545

Db 2028 HQPSFALNRRTGDG-----IVQKLSQADSGFLKEXKLGLNQRDALVALVYKORQPLKGS 2083

QY 1546 SE-----OQALHTVOYGLIKLISKTALRHFTPDV----- 1576

Db 2084 KQVMKRYRHQJDBEITCMLTAEHAKORSTTELGEMIOELRDLTQEMEVAEKLKWLNRT 2143

QY 1577 COLLDSQDLAEYNFLFALSTFTPTPDESEVAPSGFTLLATVVALNML 1625

Db 2144 LEMLDKSLSPERD-----KISES-----LRIVNMWTKI 2174

RESULT 22

US-08-265-967C-2

/ Sequence 2, Application US/08265967C

/ Patent No. 6476200

/ GENERAL INFORMATION:

/ APPLICANT: SABATINI, DAVID M.

/ APPLICANT: ERDJUMENT-BROMAGE, HEDIE

/ APPLICANT: LOI, MARY

/ APPLICANT: TEMPEST, PAUL

/ APPLICANT: SNYDER, SOLOMON H.

/ TITLE OF INVENTION: MAMMALIAN PROTEINS THAT BIND TO FKBP12

/ NUMBER OF SEQUENCES: 14

/ CORRESPONDENCE ADDRESS:

/ ADDRESSEE: BANNER & ALLEGRETTI, LTD

/ STREET: 1001 G STREET, N.W., 11TH FLOOR

/ CITY: WASHINGTON

/ STATE: D.C.

/ COUNTRY: U.S.A.

/ ZIP: 20001-4597

/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk

/ COMPUTER: IBM PC compatible

/ OPERATING SYSTEM: PC-DOS/MS-DOS

/ SOFTWARE: Patentin Release #1.0, Version #1.25

/ CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/08/265, 967C

/ FILING DATE: 27-JUN-1994

/ CLASSIFICATION: 435

/ ATTORNEY/AGENT INFORMATION:

/ NAME: KAGAN, SARAH A.

/ REGISTRATION NUMBER: 32,141

/ REFERENCE/DOCKET NUMBER: 01107,46363

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: 202-508-9100

/ TELEFAX: 202-508-9239

/ TELEX: 197430 BBMB UT

/ INFORMATION FOR SEQ ID NO: 2:

/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 2470 amino acids

/ TYPE: amino acid

/ TOPOLOGY: linear

/ MOLECULE TYPE: protein

/ ORIGINAL SOURCE:

/ ORGANISM: Saccharomyces cerevisiae

US-08-265-967C-2

Query Match 1.4%; Score 127; DB 4; Length 2470;

Best Local Similarity 18.6%; Pred. No. 0.089;

Matches 322; Conservative 236; Mismatches 597; Indels 574; Gaps 86;

QY 160 VEVADCVDKLEKELVSKYRQO-----FEELYKTEAPMTWTHGNLMTRO-----VSR 206

Db 68 VNFPRILEKIFRELTSDYKBERKLASISLFDLVSL-----HELSIEEFQAIISNDINN 121

QY 207 WFOCLAEQS-----MLEIIFLYAYFEMAPSDLLVLTVMKEQSGFSRQTNHLV 258

Db 122 KIELVHTKKTNRVGVNLISIDTLISFYAYTEELFN-----ETSR--- 161

QY 259 DETMDPFVDRIGYPSALILVEGMDI-----ESHKCALDDRRELHQFADGLICQMDCL 313

Db 162 -----LAGYLRGLIPSNDEVMRLAKTKGLKLAVPG-----GYTSDPVEFEIKSCL 208

QY 314 -MLTFG-----DIPHAAPVLLAWALLRHT--LNPETS SVVRKIGT-----A 353

Db 209 EMLTASTEKNSFSSSKPDHAKHALLITLAEKCPYLQVYNSLIDNWRALRDPHLV 268

QY 354 IQANVFOYLRLLQSLASGNDCTTS-----TACMCYGLISFVLTSIE-LH-TLGNQODI 407

Db 269 IRIDASITLAKCLSTLNRDPQLTSQVORLASCCEVG---FQVNTLECHHALLVYKEI 325

QY 408 IDTFACEVLADPSLPELFWGTEPTSGILIDSVC-----GMFHLSPLLQLLR 456

Db 326 L-----FLKDPFLNQVF-----DMCLNCIAVENHAKAKIREKIYQIVPLA 367

QY 457 ALVSGSTAKKVSFLDKSFLNELYKHKPHDIVSHEDGTLWRQTPKLLYPLGGQTNLR 516

Db 368 SFNPOLFAGKYLHQIMDN---YLEILTNAPAKKIPH-----LKDRKQLILISIG--- 413

QY 517 IPQGTQVWLDDRAVLVREYSSWTLFTCEIEMLVHVSADVIQHQRYKPIIDL 576

Db 414 -----DIATFV-----GPDIAFY---VKQILDYI 434

QY 577 HKVISTDLASIDCLPITSRIYMLQRLTVISPP-----VDVIASC-----VNC 621

Db 435 EHDLOTRKFKFRK---KFENEIFYCIGRLAVPGVGLKLNRIIDLMPKCPISDYMGET 491

QY 622 LVTLANRPA-----KWTDLRHTGFL-PRVAPVSLSGMTAEGNAGCYGN 669

Db 492 FOILTERIPSLGKINDELINLVCSTLSTGTPFLOPSPMEIPFSFERARERMRNK---S 547

QY 670 LLM---NSEPOGEYGVTIAPLRLITLVKGOIGSTOGSLVPCVWFVLEKMLPSYHKMR 726

Db 548 ILQKTESUNDNDNDIKIIOAFRLMK-N-ISKRP-----SLVEFRIVALSYI-----E 594

QY 727 YNSHGVREOIGCLIELHAILNLCHETDLSSHTPSLOFLCISLAYTEAGQVYINING 786

Db 595 HTDPRVRKLAULTSCS-IYKONI CKQTSLSLNTVS-----EVLSKILA 638

QY 787, IGVDTIDMMAAPRSRDAEGGQGGQLIKYKLAFSVTNNYIRLKPSPNVVSPLEQALS 846

Db 639 ITI-----ADPLQD-----IRLE-----VLKNLPCCFD 661

QY 847 OHGAGNNLIAYLAKIYHKHPA--LPLAIOQLRLATVAPMSVYACLGNDAAIRDA 904

Db 662 POLAOPDNRLPLFTL---HDSFNIGVAMIEIVGRLLSVPAVYIPSI---KRLILE 713

QY 905 FILRLQ---SKIEDRIKMIIEFLTVAVETOPGLIEFLNEVKDGSQSK----- 953

Db 714 LITKLKFTSSREKETSLCTLRSSKDVAKPYIEPLNVLPRKQTSSTVASTAIR 773

QY 954 ---EFSL---GWSCLHVLLEIDSQQODR-----YMCPELH 985

Db 774 TIGELSVGGEDMKIYKDLFPLIKTRPODSNSFKREBAALKALQOLAASSGYVDPLD 833

QY 986 RAALVFLHALMODRDSAMLVLRPKFEMNL---TSPFGTLS--PSETSEPSILET- 1039

Db 834 YPELL-----GILVNIILKTEMSQNRIRQTVLIGLIGLIDPYRQKEREVTST 881

QY 1040 -----CALIMKI---CLEIYY---VYKSLDSQDKTLLKKSIEKRFAYMGVY 1083

Db 882 DISTQONAPDIDALLMGMSPSNDEYTTVIVHCLIKLMD-----PSLSYHTAVI 934

QY 1084 KSLAVVAETEGSSCTSLSE-----YQMLVSAWRMLLIATTHA 1122

Db 935 QAI-WHIFTOTGLKCVSFLDQIIPITLDVWRTCSQSLLEFFQOLCS---LITIVRH- 988



```

OY 1123 DIMHLSBSVNRQLPDLVDGFKALLVPAVSNCI-----RLSGMKCTJLLILRLQ 1173
Db 989 -IRPHVDSIF--QAIKDPSSVAKQITLVSVEALSKALBEGEFKVLPTLLTLEFVILEN 1045
OY 1174 WKRELGSVEILGPYTEILEGVLAGDQOLMEKTKAKVESAFITVLQMKEMKSDIPQYSQ 1233
Db 1046 -----DSSSKVLSR--RVLRLLSEPGNLEGSH 1073
OY 1234 L-----VLNVCE---TLQGEVIALFDQTRHSLALGASATEDKDSMETDCCSRHRDQDG 1285
Db 1074 LITPKIVQMAEFTSGNLRSAI-----ITIGKLAKVDVLFEMS--SRIVHSLR-- 1120
OY 1286 VCVLGHMLAKELCEVDEGDSDSLQYTR-----LPIPLTLTLTVEVLSRKQNHFT-- 1337
Db 1121 --VLSSTTSDELSKYIMNTLSLLIQMGTSFAIFIPVINEVL-----MKGHIOHTY 1170
OY 1338 -EATLHL-----LTLARTQOGAT-----VAGAGITOSICPL-LSVYQLSNTPAQ 1383
Db 1171 DDLTRILANVDLPRKIEAHTTVDPRQEMAAAGAK--LPINSVLSKANSSSQ 1222
OY 1384 TPSASRKSGLDAPSWGYYRLSWSLMEOLIKTLRYNFLPEALDFVGHQERTLOCLNAVRT 1443
Db 1228 -----RTKEDQEW-----SKRLSIQLTKESPSHALBACSNLWAMYPLAKEFNT--- 1273
OY 1444 VQSLACLEADHTVGFILQLSFMKEMFHLBOLMRDQVNLGVCOACTSILHSRKLQ 1503
Db 1274 --AFACV-----WTELXSQYQEDL---IGSICILSPLNPEHTQ 1305
OY 1504 HYL-----QNKNGDGLP---SAVAVORYRPPESAASAPSSKOPAADTEASPOALHTVQ 1555
Db 1310 TLNLVEMEHDDKALPIPTQSLGEYAEBCAAYALAHYKEIKFTKEPBNSTIESLISIN 1365
OY 1556 YGLKLILSKTLAA--LNFHTPPDVCOILLDOSLDLAEYNF-----LPLASTPTPTF 1603
Db 1370 ----NQLNQTDAIGILKHAQG-----HSLQKTEWFEKLERMEDALHAYNEREKAG 1418
OY 1604 DSEVAPSGTLLATVVALNMLGELDKKKEPLTQAVGLSTQAEGRITLK 1652
Db 1419 DTSVSVTLGKKR-----SLHALGEW-----EQLSOLAARKWVSKLQTK 1458

RESULT 23
US-08-305-790B-3
/ Sequence 3, Application US/08305790B
/ Patent No. 6492106
/ GENERAL INFORMATION:
/ APPLICANT: SABATINI, DAVID M.
/ APPLICANT: ERDUMENT-BROMAGE, HEDIEE
/ APPLICANT: LUI, MARY
/ APPLICANT: TEMPEST, PAUL
/ APPLICANT: SNYDER, SOLOMON H.
/ TITLE OF INVENTION: MAMMALIAN PROTEINS THAT BIND TO FKBP12
/ NUMBER OF SEQUENCES: 15
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: BANNER & ALLEGRETTI, LTD
/ STREET: 1001 G STREET, N.W., 11TH FLOOR
/ CITY: WASHINGTON
/ STATE: D.C.
/ COUNTRY: U.S.A.
/ ZIP: 20001-4597
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/305,790B
/ FILING DATE:
/ CLASSIFICATION: 536
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/265,967
/ FILING DATE: 27-JUN-1994
/

```

[illegible]

```

QY      847 OHGAHGNLLIAVLAKTYHKKDPA--LPRLAIQLKRLATVAPMSVYACLGNDAAAIRDA 904
      662 POLAQNDRNLRLFTAL-----HDSFNIQSVAMELVRLSSVNPAYTPTI-----RKILIE 713
QY      905 FLTRLO--SKIEDMRKMIIEFLTVAVETOPGLIEFLNTEVKDGDGSK-----953
      714 LITKLFSTSSREKETASILCTLRSSKQVAKPYIEPLNLVLPFGQDTSSTVASTLR 773
QY      954 ---EFSL---GWMSCILHVLLEIDSOQDR-----YKCPPLIH 985
      774 TIGELSVVGEDMKIYKODLPFLITTFODQNSFRERALKALQLAASGCVIDPLID 833
QY      986 RAIAIHALMODRDSAMLVLRKPKFWENT--TSPLEGLIS--PSESTSEPSILET- 1039
      834 YPELL-----GLVNIKTEGONIRQVTVLIGLIGALIDPRCKEREVSTTY 881
QY      1040 -----CALMKII--CLEIYY--VYKGLSDGLKDTLKKFKSTKRFAYMSGYV 1083
      882 DISTEQNAPRIDIALMOGMSPSNDEYTVTVIHCLIKLKD-----PSLSYHTAVI 934
QY      1084 KSLAVVAETEGSSCTSLF-----YOMLVSAMRMILLIATTHA 1122
      935 QAI-WEIFPTGLKCVSFLDOIPTILDMVKTCSQSLSEFYIQLCS---LITIVRQH- 988
QY      1123 DIMALTDVVRQQLFDVLDTGKALLVPAVNCU-----NLGSMKCTLLILILRQ 1173
      989 -IRPHVDSIF--QAIKDFESSVAKLQITIVSVIEAISKALEGFEKRLVPLTILFLVILEN 1045
QY      1174 WRRELGVDELIGPTEILEGLVLOADQOLMEKTKAKVPSAFITVQMKEMKVSIDIPOYSQ 1233
      1046 -----DKSDKVLISR--RVLRILLESFGPNLEGYSH 1073
QY      1234 L--VLNVC-----TLQEVYALPDQTHSLATGATEDKXDMETDSCSRHRDQDG 1285
      1074 LITPKIVQMAEFTSGNLRSAI-----ITIGKLAKVDLPEMS--SIVIVSLAR-- 1120
QY      1286 VCVLGLHLAKELCEYDEGDSWLOVTR-----LPILPTLLTLEVSLRKONLHFT- 1337
      1121 -VLSSTTSDELISKVIMNTLSLLIOMGTSFAIFPVIENEVL-----MKKHIOHTIY 1170
QY      1338 -EATLHL-----LTLARTQOGATA-----VAGAGITQSICPL--LSVYQDSTNGTAQ 1383
      1171 DDLTRNLINNDVLPKILEANTTTPKPAQOMHADAGVAK--LPINOSVLKSMANSQO 1227
QY      1384 TPBSAKSLDAPSWGVTRLSMSLEOULKTLRYNLEALDFVGVQERTLOCLINVRT 1443
      1228 ---RTKEDWQEW-----SKRLSIQILKESPSHALACSNLASMYVPLAKELFNT- 1273
QY      1444 VQSLACEADHTVGFILQSNFMKEWHFHLPOLMRDIOVNLGYLCOACTSLHSRKMLO 1503
      1274 --ARACV-----WTELYSQOEDL--IGSLCTALSSPLNPPRIHQ 1309
QY      1504 HYL-----QNKNGDGLP--SAVAQRYQRPSPASAAAPSSSKQPAAPTEASEQOALHTVQ 1555
      1310 TLLNLVEFMERHDKALPIPTQSLGEVAMERCHAYAKALHYKEIKIKEPENSTIESLSIN 1369
QY      1556 YGLKLISKTLAA--LRHFTPDVCQIILDOSLDLAEYNF-----LPAISFTPTTF 1603
      1370 ---NQNLQDTPAAGIILKHAQO-----HNSLOLKETWEKLERMEDALHAYVEREKAG 1418
QY      1604 DSEVAPSGFTLLATVVALNMLGELDKKEPLTQAVGLSTQABGSTRLLK 1632
      1419 DTSVSVTLGKMR-----SIHALGEM--EQLSOLAAKMKVKSGLQTYKK 1458

```

RESULT 24

US-09-248-796A-15028

Sequence 15028, Application US/09248796A

Patent No. 6747137

GENERAL INFORMATION:

APPLICANT: Keith Weinstock et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

```

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248, 796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074, 725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096, 409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 15028
; LENGTH: 1106
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-15028

Query Match      1.4%; Score 126.5; DB 4; Length 1106;
Best Local Similarity 18.4%; Pred. No. 0.022;
Matches 230; Conservative 182; Mismatches 396; Indels 441; Gaps 65;

QY      382 CMCVYGLSLFVL-----TSLHLTLGNQODI-IDTACEVLADPSLPE 422
      11 CVCLVQNHFEVLFPDSQYSYTSLSFSKATNPPEVLVQWMLYIFACSTLTATNDOK 70
QY      423 LFMGTBFTSGIITLDSVCGMFPHLSPILLQL--RALVSGSKTAKYVSFLDKMSFYN 479
      71 ISPPNQPTH-----RKSLQWTFIOHOKITSKSVKMWPLPL-KLS--Q 112
QY      480 ELYKHPHVDYSHEDGTMRRQTPKLLYPGQ--TNL--RIPQSTGVQMLDRAVYV 534
      113 PMVESVYGLSHININILKSFLENI--PLAVEWNTNKTNPQ-----DDAFRI 161
QY      535 RMESYSWTLFCEJEMLHVSTADVIOHQRVAPIIDL-----YHKVISTDLSTIAC 589
      162 E-----VHHIT-----NLTEKGLHLLITDSDIVAN 189
QY      590 LPTSRITWMLQRLTVISPPDVYASC-----VNCITVLAARNPAKWTDLR 638
      190 LVAIVKVNKFLS-----EPALQTFECQRLRRYFCGLENVYGLKSDLKK- 239
QY      639 HNGEFLFVAPPS-----SLQMSAAGMN-----AGYGNILANSQPO 678
      240 ---LPEFARVACENYKEMCGYGSAAVTERYNAIIRIKTERKAGPSALAILERKX 295
QY      679 GEYGVITAFRLITLVKQGSTOSQGLVPCMVFLKEMLSYHKM-----RYNSHGV 733
      296 ---LQSLAKGMAILCSQI--KKKLEIPGNLAVSVDIGLMMNTQALLSEVETIK 348
QY      734 EQIGCL-----LELHAILNLCH--ETDHS-----SHTPSLOFL 767
      349 EIGHCSRNVIENKFEFODIISAVVORCVSQTSLSEYCLFVEIFEMKNTFEIYSDILV 408
QY      768 CICS-LAYTEAGQVININGIGVDITIDVWMAOPRSDAGEQ-----GGQGLIKTV 818
      409 CLASVLTANESPO-----VAKSAIILIKYFEKSLGSESLERYTEAICSSSVYK- 459
QY      819 KLAFSVTNNVIRLKPS--NVVSPLEQALSOHGAHGNLI-----AVLAKYI 863
      460 KYLFDLSNDVQNIKESLYNRISYLTQNI-----NLIPPAAREALISLVNWSKYT 511
QY      864 --YHK-----HPALPRLAIQLKRLATV-----APMS-----VYACLGNDAA 900
      512 LNYETPLPEKOSTDIHEKAPDSSSLMWLSNLFELTVFGSSMSQVEVLWISNNVON 571
QY      901 IRDAFLTRLOSKIE-----DMRIKMIIEFLTVAVETOPGLIEFLN-----LE 944
      572 IDKIYVFLQSSLELAKNSFVQSGRQIIDYLAFAPSDPAIYIIDKILNNLQRPMMVPPOLG 631
QY      945 VQSGDSGSK-----FSLGMS-----CLH 964
      632 NKSTAPISKDYPVYCDIRKIITQNEKQVIFSLQSLMWFLVDIYTLKNDMLTEKPLIHL 691
QY      965 AVLEILDSQOQDRWCPLHRAA--IAFLHALMOD--RDSAMLVLR-----TKRFTW 1014

```

Db 692 VSLAMD-----HYLPIVBOAGTFLVHLIHALVRDDPKAQETILSLRENDHITHLVNY 745  
Qy 1015 ENLTSPLFTGLSPB-----ETSEPSILE-----TCALIMKIC--L 1049  
Db 746 DDLNNDKKGVTPPKMDDLIRSAIKMFAVAVNIQDDMSRVSLHWATICA--VRHJACRSF 804  
Qy 1050 EIVYVVKSLPOS--LQDTLKFS--IEKRFVWSGVYSLAVH--AETEGSSCTSLLEYQ 1105  
Db 805 QVPSLSLFLDOSMLKMDLHRISNTISQETVDIQSFAMQIMLTNAITAEIDSECLIPF 864  
Qy 1106 MVSAMRMLLIATTHADIMHLTDSVVRQLFDVLDG-----TKALLVPASVNC-- 1157  
Db 865 QLF--WCVACSLSTH-----EQEPIEVLSTLTKFKISKIDDAADVAQLKXT 910  
Qy 1158 -----RLGSKCTLLILLR--QMKRELSVDEILGPIITELIEGVLDQDLMEKTK 1207  
Db 911 PPOKMGFEGIQ-QVILVGRSSSTSWPTLTKFLDKL-----IVHXSLSLIGMD 959  
Qy 1208 AKVPSAFITVLOMKEMKVSIPQVQLVLANVCELTQEEVIALFDQTRHSLALGATEDKD 1267  
Db 960 SRVLTALL-----ANMPRF-----LHLSSENVTPETENT--GMALSQLAEDSG 1001  
Qy 1268 SMETDDCSRRHRDQDGVCVLGLHLAKELCEVEDGDSMWLQVTRRLPIPLTLTLEVS 1327  
Db 1002 KSS-----LAKLIVSPDK-----RR--FRSKDPLDQS 1027  
Qy 1328 LRKONLHFTF---ATHLTLTANTQOGATNAVAGITQISICLPLLSV 1373  
Db 1028 ISCIKNAPFPEYOALVLLGLLFLNKMVPWKTETLGLQLI--FPLVDL 1075

## RESULT 25

US-09-085-199B-9  
; Sequence 9, Application US/09085199B  
; Patent No. 6235879  
; GENERAL INFORMATION:  
; APPLICANT: Hayden, Michael R.  
; APPLICANT: Hackam, Abigail  
; APPLICANT: Huq, A.H.M. Mahbubul  
; APPLICANT: Chopra, Vikramjit Singh  
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the  
; TITLE OF INVENTION: Huntington's Disease Gene  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Opedahl & Larson  
; STREET: PO Box 5270  
; CITY: Frisco  
; STATE: CO  
; COUNTRY: USA  
; ZIP: 80443-5270  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 KB storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS DOS 5.0  
; SOFTWARE: WordPerfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/085.199B  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Larson, Marina T.  
; REGISTRATION NUMBER: 32038  
; REFERENCE/DOCKET NUMBER: UBC-P-013US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (970) 668-2050  
; TELEFAX: (970) 668-2052  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 756  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

; HYPOTHETICAL: no  
; ORIGINAL SOURCE:  
; ORGANISM: mouse  
; FEATURE:  
; OTHER INFORMATION: Huntington-interacting protein  
US-09-085-199B-9

Query Match 1.4%; Score 125.5; DB 3; Length 756;  
Best Local Similarity 20.2%; Pred. No. 0.014; Gaps 26;  
Matches 136; Conservative 96; Mismatches 261; Indels 181; Gaps 26;

Qy 1126 HLTVSVVR-----QFLVDLDTGKALLVPASVNC-----RLGSKCTLL 1168  
Db 94 HLIERLVREISGLQGLDNMKIESQAMLDQKGVSELEHARQHLGRAMDCELR 153  
Qy 1169 ILRQMKRELSVDEILGPIITELIEGVLDQDLMEKTKAVPSAFITVLOMKEMKVS 1228  
Db 154 TELDELKQREDTEKARSLTEI--ERKAQANEQRYSKLKEK----- 193  
Qy 1229 PQVSQLVLANVCELTQEEVIALFDQTRHSLALGATEDKDSMET-----DDCSRRHRDQ 1262  
Db 194 --YSELVQNHADILRKNAEVTKQVSVARQA-QVLEBEKELADSPARVSDQQRKTOEQ 250  
Qy 1283 RDGVCVGLHLAKELCEVD-----EDGDSMWLQVTRRLPIPLTLTLEVSLRMKQ 1332  
Db 251 QDVLENLKHELATSRQELQVHLSNLETSAGSEAKML-----TQIAELE--KEG 297  
Qy 1333 NLHFTFATLHLTLTANTQOGATNAVAGITQISICLPLLSVYQSLSTNGTAQTPSARKSL 1392  
Db 298 SLATVAQREBELALRDQLESTQIKLAGQESWC-----QQVDRQKRTL 342  
Qy 1393 DAPSMQGVYRLSMELQGLKTLRYNPLRELDVGVNHOERTLQCLNANRTVQSLCLUE 1452  
Db 343 LA-----GIRKAEREIQEALSQ-----EEPTL-----ISCAGS 372  
Qy 1453 ADHTVGFITQLSNPMKEWHFPLPOLMRDIQVNLGCLQAC---TSLHSRKMLOHYLON 1508  
Db 373 TDHLSKVS SVSCLBQLEKKGSG-----YL--ACGEDISELHSTLLAHL--- 417  
Qy 1509 KNGGCLPSANQVQRPSPASAAAPSSKQPADTER--SEQQLHYVQGLKILSKTL 1566  
Db 418 -TGQTVIQGSATSRAPPEPADSLTEACROYGRETLAVLSLEEGVENADVALRNCL 476  
Qy 1567 AALRHFTPDVQQLLDQSLDAEYNPLFALSFTPTDSEVAPSPGTLATVVALNMG 1626  
Db 477 SRVTLGEE---LPPGLDKQEL-----GDLVDKEMAY---SAILEATTRIE 521  
Qy 1627 ELDKKKEPTQAVGLSTQAEGRTRILKSLMFTMENCEFLLISQMRYLDRPAVHRDQR 1686  
Db 522 EILSKSRAQDGVGLVNNERTILGCTSLM-----QAIKVL---VVASKOLQK 565  
Qy 1687 MKQEL---SSELSTLLSLSHYFRGAPSSPATG-----VLPSFGKSTSLSKA 1732  
Db 566 EIVSGRGSAPKCFYANSRHTEGLISASRAVGMGATIMVADADLVQGGKKEELMVC 625  
Qy 1733 SPESQEPILQVOA 1746  
Db 626 SREIAASTAQLVAA 639

## RESULT 26

US-08-726-214-8  
; Sequence 8, Application US/08726214  
; Patent No. 6107076  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Wei-Jen  
; APPLICANT: Gilman, Alfred G.  
; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE  
; TITLE OF INVENTION: AND USES THEREFOR  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433

CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/726,214  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/005,498  
FILING DATE: 04-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 37,642  
REFERENCE/DOCKET NUMBER: UTSD:450  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 474-7577  
TELEFAX: (512) 418-3000  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1064 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-726-214-8

Query Match 1.4%; Score 125.5; DB 3; Length 1064;  
Best Local Similarity 19.4%; Pred. No. 0.026;  
Matches 115; Conservative 74; Mismatches 166; Indels 227; Gaps 26;

QY 966 VLELDSDQOORVYCPPLHRAAIFLALMODRDSAMLVKTRPK----- 1012  
DB 541 VIBOLNSQKQ-----WKQSKDFLLTLTYREKEMKQYRLSALPA 580  
QY 1013 -----FWENLTSPLEFGLTSPSESTSPSTLEFCAIMKICLEY-----VYKGS 1058  
DB 581 FKYYACFTFVFLFSLFTQIMLVTRPPALATTYSTIFLLFLFVCSSEHLTCVQKG- 639  
QY 1059 LDQSLKDTLKRFSEKRPAYSGVYKSLAVHAETEG-----SSCTSLLEYQMLVSAMRM 1113  
DB 640 -----PKMLHW-----LPALSVLVATRPGLRVALGTATILVFTMAV-----V 677  
QY 1114 LLITATTHADIMHLDSVVRRLPLD-VLDGTAKLLLVPAVNCRLRGSMTCTL----- 1166  
DB 678 SLFLFVSSDCPFLAPNVSSVAENVTSWELPASLPILISTPYSMHCCVGLFSLCSLFLMSF 737  
QY 1167 ----- 1166  
DB 738 ELKLLLLMLVASCSLFISHAWMLSDCLLARLYOGSLGSRPVYKPKLMGATYFIFF 797  
QY 1167 -----LLILLRQ-----WKRLGSVDIELGPLTEILEGVLOA-----DQQL 1202  
DB 798 FTLLVLAKQNEYCYRDLFLMKKRLQRE-----ETETMENTLPAHVAPOULIGQNRREDEL 853  
QY 1203 MEKTKAKVFAFITYLQWKEKVKSDIPQYSQVLN-----VCETLQEBVIALFPD----- 1252  
DB 854 YHGVCEVCVLPFAIIPPEK-----YSSNNHGELEGLRLINELIADFELLSKPK 906  
QY 1253 -----TRNSLAIG-SATEDKDSMETDSCRS-RHNDORDGVCV-LGLHLAKELC 1298  
DB 907 FSGVEKIKTIGSTYMAATGLNATPGODTQ--DAERSCSHL-----GTMVEEAVALGSKLG 960  
QY 1299 EVNDEDSQSMQVTRRL--PILFTLL-----TTLEVSIRMQQ-----NLHTEA 1339  
DB 961 VINHGSFNNFELRLKVLHNGVPVAVGIAQKQFYDIMGWNTVAVASRMSTGVLRKIQVTEE 1020  
QY 1340 TLHLTLTART--OOGATAVAGAGITQSLCLPLLSVYQLSTNGR-AQTPSAS 1388

DB 1021 TARALQSLGTCYRGVIRKYGKG-----QLCTYFLMTDLTRTGSPPAS 1064

RESULT 27  
US-09-085-199B-4  
Sequence 4, Application US/09085199B  
Patent No. 6235879  
GENERAL INFORMATION:  
APPLICANT: Hayden, Michael R.  
APPLICANT: Hackam, Abigail  
APPLICANT: Huq, A.H.M. Mahbubul  
APPLICANT: Chopra, Vikramjit Singh  
APPLICANT: Kalchman, Michael  
TITLE OF INVENTION: Apoptosis Modulators That Interact with the  
NUMBER OF INVENTIONS: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Opedahl & Larson  
STREET: PO Box 5270  
CITY: Frisco  
STATE: CO  
COUNTRY: USA  
ZIP: 80443-5270  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS DOS 5.0  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/085,199B  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Larson, Marina T.  
REGISTRATION NUMBER: 32038  
REFERENCE/DOCKET NUMBER: UBC.P-013052  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (970) 668-2050  
TELEFAX: (970) 668-2052  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 914  
TYPE: amino acid  
MOLECULE TYPE: protein  
TOPOLOGY: linear  
HYPOTHEICAL: no  
ORIGINAL SOURCE:  
ORGANISM: human  
FEATURE: Huntington-interacting protein  
US-09-085-199B-4

Query Match 1.4%; Score 125; DB 3; Length 914;  
Best Local Similarity 21.6%; Pred. No. 0.022;  
Matches 134; Conservative 76; Mismatches 244; Indels 166; Gaps 27;

QY 1164 CTLLILRLQWRELSQVDEILGPLTEILEGVLOAQDQIMETKAKVFAFITYLQWKEM 1223  
DB 307 CEFLRAELDELRLQRREDTEKAQRSLSL-ERRAQANEGKYSLTKK----- 351  
QY 1224 KVSIDIQYQVLYNVETIQE--EYIALPDQTRHSIALGASATEDKDSMETDSCSRHND 1281  
DB 352 -----YSELVQNHADLLRKNAEYTKQVSMARQA-----QVDLREKKELED-SLERISD 399  
QY 1282 QRDGVCVLGLHLAKELCEVEDGDSWLQVTRRLPILPTLLTTLVSLRMKQVLFHTAATL 1341  
DB 400 Q-----GQRTQDELEVLDESLSKQELGTSQR--ELQVLOGLSLETSAQSEANMAAFAPAL 450  
QY 1342 HLTLTLTARTQAGATAVAGAGITQSLCLPLLSVYQLSTNGTACTPSASRSKSLDAPSPGVY 1401  
DB 451 E-----KERDSVYSGAARE-----BELSALRRELDGT----- 478  
QY 1402 RLMSLME-----QLKTLRYNLPALDPYGVHQ--ERTLQ--CLNAARVQSLACLEAD 1454

Db 479 QLKLASTEESMCQLAKDQRKML-----VGSRKAAEQVLDALNQLPEPPLISGAGSAD 532  
Qy 1455 HTVGIILQSNFM-----KEMHPLPOLMRDIOVNLGYLQAC-----TSLHNRKRLQHYL 1506  
Db 533 HLLSTVTSISSGIEQLERKSWQYL-----ACPDIGSLHSTITLALH- 575  
Qy 1507 QNKINQDGLPSAVALGVRPPSAAAPSSKOPADTEA-----SEOOALHTVQGLK- 1560  
Db 576 ---TSDALAHGATTCRLRAPPEPADSLTEACKQYGETLAIYLALESBSLENAADSTAKRN 632  
Qy 1561 ILSKTLAALRHFTPDVQCILLDQSDLAELYNFLFALSFTPTPDESVAPSGTLLATVNV 1620  
Db 633 CLSKIKA-----IGEBLLPRGLDIKQEBL-----GDLVDKEMAT-----SALET 673  
Qy 1621 ALNMGELDKKEPELTQAVGSTQAEGRITLKSILMTFMENCFYLLISQAMKYLDPAYH 1680  
Db 674 CTARLEEMLSKSRACDGTGKLENNERILRCTSLM-----QALQVL---IYA 717  
Qy 1681 PRDKORMKOE-----LSSELSTLLSLRYFRGAP--SSPATGVLPSPQGS 1726  
Db 718 SKDLQREIVESGRGTASPRKEFYAKNSRMTBGLISASKAVGATVMDAADLVVGRKGF 777  
Qy 1727 TSLSKASPSGDEPLIQLVQA 1746  
Db 778 EELMVCSEHIAASTAQLVAA 797

RESULT 28  
US-09-228-246-2  
; Sequence 2, Application US/09228246  
; Patent No. 6245510  
; GENERAL INFORMATION:  
; APPLICANT: Staerkawicz, B. S. et al.  
; TITLE OF INVENTION: PRP Protein and Nucleic Acid Sequences: Compositions  
; and Methods for Plant Pathogen Resistance  
; FILE REFERENCE: 51700  
; CURRENT APPLICATION NUMBER: US/09/228,246  
; CURRENT FILING DATE: 1999-01-11  
; EARLIER APPLICATION NUMBER: 08/680,327  
; EARLIER FILING DATE: 1996-07-11  
; EARLIER APPLICATION NUMBER: 08/310,912  
; EARLIER FILING DATE: 1994-09-22  
; EARLIER APPLICATION NUMBER: 08/227,360  
; EARLIER FILING DATE: 1994-04-13  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1824  
; TYPE: PRP  
; ORGANISM: Lycopersicon esculentum  
US-09-228-246-2

Query Match 1.4%; Score 124.5; DB 3; Length 1824;  
Best Local Similarity 19.0%; Pred. No. 0.09;  
Matches 305; Conservative 209; Mismatches 552; Indels 541; Gaps 81;

Qy 40 LNKMKRLLLEGLSYKPPSP-----SSAEKVYKANK---DVASPLK-ELGL 80  
Db 214 MODESKFLERIS--RPSFPGRLQVYALNALIGLNSRSKTTNNAKYMDFYSALODDLRL 271  
Qy 81 RISKFLGDEBOSVOLLOCYLOEDYRGTDSVKTYLODEROSQALILKAD----- 131  
Db 272 RC-----DNRIMLQRLGSLYLCRFLRD-IESIPVSHRQILSIQLNMEDLAISANA 322  
Qy 132 YYEERTCILRCVLLHLYTFQDERHPYRV-----YADCVDKLEKEIV- 174  
Db 323 YSYEDMDKTSIDHDLPHLQMKFNVVKEVDLIRLQNIQTIIVPMKDLIDYVMEELMF 382  
Qy 175 ---SKTROQPEF-----LYKTEAPTWETHGNL---MTERQVSRMNVQCLREOS 216  
Db 383 FRSYMDAFDQKEQTRITVILNYIOSAVSQAMSVCDSDCHDLNQNDLAR-EINCLHFO- 440  
Qy 217 MLLEIFLYYAVAFENAPS-----DLVLVTKPFKEQ-----GFGSRQ 252

Db 441 LLLKFKFKVALRQKCPISASSTPDHDMIDLNLFPNNFEALIDSSYSSMLKASCPSSHR 500  
Qy 253 TNRHLVDETMD---PVDRIQYPSALILVBGM-ITSLHKAL-----DDRR----- 295  
Db 501 PNRDAESPNTSFLCGPNTDVVSFYSSSSRIPMDILKRFHEYIYVNLIRKXETMLTFTI 560  
Qy 296 ---ELHQFQDGLIQODMCMILTF---GDIPHNAVLLAMLLRHTLAPNESSVY----- 346  
Db 561 ADEVKRF-YDGLL-----LMVYLIPEVPHTBCKRQNDLSMRHEAVALIESAVCHAY 613  
Qy 347 -RKIGFTAIQNL-VFQYLTRLLQSLASGANDCTTSTACCVYGLSFVYLSLEHL--TLG 402  
Db 614 EDMNNNRREINQVQLFTVTFWLIKSGN-----LMDLLKHKSTLG 655  
Qy 403 NO-QDIITPAC-----VLAD-----PSLPELPKGTPTSGGILLDSVC-----G 442  
Db 656 NOVLDLIESAHBELILNLSIMDLRKLYRLDLMHABVYAKRLAIFSGCYEYFNNG 715  
Qy 443 MEPHLSPL-QLRALVSGSTAKV-----YSPFDKMSFYNELYHK-----PHDV 489  
Db 716 STETMRPLSDFLQIESVKEFRNVCLQVLDISPSFLTDBEGLVNELKMQALVPND 775  
Qy 490 ISHEDGTLM-RROTPKLLP---LGQTNLRIPQ-----TVGQ----- 524  
Db 776 AVSSDGSLEDASTETKMLPSDFLEIESVEIKERAKLYDQVLDATHGCTSKDGKSPIN 835  
Qy 525 VMLDDRVLVREWSYSSWTLFTCEI--EMLHVVS-TADVIQHCQRYKPIIDLVHAYIS 581  
Db 836 IMLTQODKLPDYDASVSYLLNQISVVDKLIHGISLVLDIVQYNNHIEILDLEAVOD 895  
Qy 582 TDLSDADLLPITSNI---VWL-----LQRLTVISPPVDVYASVNCGLTAAANPAKV 633  
Db 896 KNYI---CPFSVKGYIPAWYTLTYLSDVKQLKFEABEVKI---CLKV----- 938  
Qy 634 WTLDRHTGFLPFAVPVSLQMSIAEAGMAGYGNLMMN--SEOPQGEYGTIAFLRLI 691  
Db 939 -----PDSSTSYSPKTNL---GYLNCILGLLELRKLDLIDLKQI 980  
Qy 692 TTLVKGQI-----GSTQSGQVPCVM-----FVLKEMLPSTYHKRYNSH 730  
Db 981 ESYKEGILLCLNSFIDHFESEYDEHDEACGLIARVSVAKAEYVIDSCIAVSHPLMYVL 1040  
Qy 731 GVRBQIGLLIELHAILNLCHET-----DLHSHTPSLQFICISLAVTEAGQYVI 782  
Db 1041 WISE-----VLENIKLVNKVVGETCERNIEVYHEV-AKTTVYAPSPSATQANBEM 1094  
Qy 783 NIMGIGVDTI-DMVWAQPRSD-----GAEGQGGQLIKTYKLAFSVTNNVIRLRKPSN 836  
Db 1095 EGFOITIBELDKXLGSPBEDVISYVGMPLGK----- 1128  
Qy 837 VSPLEQALSGHAGHNNLIVLAKYIYHKDPLRLAIOULKRLATVAPMSVYA-CLG 895  
Db 1129 -----TLIAKKIY--NDPEV-----TSRFDVAQCVV 1153  
Qy 896 NDAALIRDAFLTRLOSKIE-----DMRITVMILEFLTVAVENQPGILEFL 941  
Db 1154 TQLYSWRELTLTLNDVLEPSDRNEKEDJADELRRLTLTRFLI-----LIDVW 1205  
Qy 942 NLEVYDG-----SDSKFSLGMMSCILHAYVELLIDSOQODRYWCPRLHRAAIAFLHLM 996  
Db 1206 DYKWDNLCMCFSDVSNRSRIILITRLNDVAVEYVCESDP-----HHLR 1249  
Qy 997 QDRDSAMLVLRKPKFMENTLSPFLGTLSPBSETSP--SILETC-----AL 1042  
Db 1250 LFRDDES-----WTLQKEVFQGESCPBELEDVGFERSKCRGLPLSVVLVAGVL 1299  
Qy 1043 IMKIICLEIYYVK-----GSLDOSLQDTIKES-----TEKRFAYSGVYKSLA 1087  
Db 1300 KQKKTTLDSMKVQESLSSQRIQSLESBI--SIIGFSYKNLPHYKPCFLTFGGGLQSKD 1357  
Qy 1088 VHVATEGSSCTSLLEYQMLVSAWPMILLIATTHADIMHLLTDSVVRQRLFVLDTGRTAL 1147

Db 1358 IHVSK-----MTKLM-----VAEGFVQANNKGEQEDTAQGLDILIGRNV 1398  
QY 1148 LIPVSVN-----CLRISGMKCTLLI-----LIRQWRELGSV-D 1182  
Db 1399 NAMEKRPNTKVTCTRIHDLHKFCHEKAKQEDPLQINSGEVFPERLEEYLFVHSYOD 1458  
QY 1183 EI-----LGPTEILEGVLOADQOIMETKAKVPSAFITVLQMKEMKSDIPQYSQVL 1236  
Db 1459 EIDLWRPSPRSNVSILFNAIDPDNLLMPRDISFIESF-----KLVKYLDLESF----- 1507  
QY 1237 NVCEFLQEEV-----IALFDQTRHSIALGASATEDXDSMETD-----DCS----- 1275  
Db 1508 NIGGFPEETIQLQMKYFAQTDANSIPSSIAKLENETTFVVRGLGEMILPCSLKMY 1567  
QY 1276 RSRHNDQRDQVCVLGLHLAKELCEVEDDGSWLQVTRRLPLPTLLT 1322  
Db 1568 KLRHIVANDRVS-FGLH-----ENMDV-LTGNSQLPNETST 1603

RESULT 29  
US-09-085-199B-5  
; Sequence 5, Application US/09085199B  
; Patent No. 6235879  
; GENERAL INFORMATION:  
; APPLICANT: Hayden, Michael R.  
; APPLICANT: Hackam, Abigail  
; APPLICANT: Hug, A.H.M. Mahubul  
; APPLICANT: Chopra, Vikramjit Singh  
; APPLICANT: Kalchman, Michael  
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the  
; TITLE OF INVENTION: Huntington's Disease Gene  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Opedahl & Larson  
; STREET: PO Box 5270  
; CITY: Ftisco  
; STATE: CO  
; COUNTRY: USA  
; ZIP: 80443-5270  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS DOS 5.0  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/085.199B  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Larson, Marina T.  
; REGISTRATION NUMBER: 32038  
; REFERENCE/DOCKET NUMBER: UBC-P-013US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (970) 668-2050  
; TELEFAX: (970) 668-2052  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1090  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: no  
; ORIGINAL SOURCE:  
; ORGANISM: human  
; FEATURE:  
; OTHER INFORMATION: Huntington-interacting protein  
; US-09-085-199B-5

Query Match 1.4%; Score 123.5; DB 3; Length 1090;  
Best Local Similarity 22.0%; Pred. No. 0.043;  
Matches 99; Conservative 61; Mismatches 164; Indels 125; Gaps 21;  
QY 1164 CTTLLILRLQKRELGSVDELIGPTEILEGVLOADQOIMETKAKVPSAFITVLQMKEM 1223

Db 483 CEFLRAELDELRRQREDTEKQSRLSSEI-EKKAQANQRYSKLKE----- 527  
QY 1224 KVSIPQYSQVLAVNCEFLQE-EVIALFDQTRHSIALGATEDKDSMETDCCSRSRHD 1281  
Db 528 -----YSELVQNHADILRRKNAEVTQVSWARQA-----QVDLREKKELED-SLERISD 575  
QY 1282 QRDGVCVLGLHLAKELCEVEDDGSWLQVTRRLPLPTLLTLEVSIMKONLHPTETL 1341  
Db 576 Q-----GQRKTEQLEULESLKQELATSOR-ELOYLGSLSTSQAQSENNMAAFPAEL 626  
QY 1342 HLLLTARQGAATVAGAGITGSIICPLLSVYQSTNGTAPTSARKSDIADSPWGVY 1401  
Db 627 E-----KERDELVSQAHRB-----EELSLARKELQDT----- 654  
QY 1402 RLMSLME-----QLKLRNLFLEALDFPGVHO-ERTLO-CLNAVTVQSLACLEAD 1454  
Db 655 QKLASTEESMCQIAKQORRKL-----VGSRRQAQVQIDALNQLSEEPPLISCAGSAD 708  
QY 1455 HTVGFIILQSNFM-----KEWHFHLPOLMRDIOVNLGYLQAC-----TSLHSRKLQHYL 1506  
Db 709 HLLSTVTSISCITQLKRSMSQYL-----ACPEIDISGLHSTITLHL- 751  
QY 1507 QNKGDLPSAVAGVORPPSASAPSSSKOPADTEA-----SECOALHTVQYGLK- 1560  
Db 752 ---TSDAIAAGATCTCLAPPEPADSLTEACKQYGRFTLAVLASLEGSGLENNADSTAMRN 808  
QY 1561 ILSTYLAALHFTDVCQIILLDQSLDLAE 1589  
Db 809 CLSKIKR-----IGEBLLPRGLDITQ 829

RESULT 30  
US-08-680-327-3  
; Sequence 3, Application US/08680327  
; Patent No. 5853521  
; GENERAL INFORMATION:  
; APPLICANT: Staekawicz, Brian S., Oldroyd, Giles Edward,  
; APPLICANT: Salmeron, John M., Rommens, Caius  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PLANT  
; TITLE OF INVENTION: PATHOGEN RESISTANCE  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klariquist Sparkman Campbell Leigh &  
; ADDRESS: Winston  
; STREET: One World Trade Center  
; STREET: 121 S.W. Salmon Street  
; STREET: Suite 1600  
; CITY: Portland  
; STATE: Oregon  
; COUNTRY: United States of America  
; ZIP: 97204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Disk, 3-1/2 inch  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/680,327  
; FILING DATE: July 11, 1996  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/310,912  
; FILING DATE: September 22, 1994  
; CLASSIFICATION: 800  
; APPLICATION NUMBER: 08/227,360  
; FILING DATE: April 13, 1994  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dow, Alan E.  
; REGISTRATION NUMBER: 35,123  
; REFERENCE/DOCKET NUMBER: 5151-45038  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (503) 226-7391  
 TELEFAX: (503) 228-9446  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1824 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-680-327-3

Query Match 14%; Score 123.5; DB 2; Length 1824;  
 Best Local Similarity 18.6%; Pred. No. 0.11;  
 Matches 300; Conservative 206; Mismatches 558; Indels 545; Gaps 77;

40 LNKHRRRLLEGISYKPPSP-----SSAEKVANK-----DVASPLK-ELGL 80  
 214 MDDESKLERIS--RPPSPGLROYVLAALGLANSRSKTYMAKTMIDPFSALODDLRL 271  
 81 RISKFLGDEBOSVOLQCYOEDYGRGTRDSVKTVALODEROSALILKIAD----- 131  
 272 RC-----DNRIWLGRLSYLCRFLRD-IESYPVHRQLISQLAMMEDAIGSANAI 322  
 132 YYBERTILACVHLITFYODERHPYVE-----YADCVDKLEKELV- 174  
 323 YSYDEDMKTSEIDHELFHLQMKFNYYVVEVDLIRLQNIQGTIIIVPMKDLIDYWEELMF 382  
 175 --SKYRQOFEE-----LYKTEAPTWETHGNL--MTEROVSRMVVOGLAREOS 216  
 383 FRSTFMADFQCKEOTRIITVILNVIQSAVSOAMVCSLCHDLNONDAR-BINCLHFO- 440  
 217 MLLEIIFLYAVFEMAPS-----DLVLVTMKFKEQ-----GFGSRQ 252  
 441 LLKRFKFKVARIQWCPISIASSTPDHIMIDLNLPMNFEALIDYSMLKASCSSSHR 500  
 253 TNRHLVDSTMD--PFVDRIGYFSALILVSGMDIESLHKCALDRRELHOPADGLICQ 308  
 501 PNRAOESPTSEFLCGPNTDVYSPYSSSRIPKMD-----ELIKFHEHIIIVNLARK 551  
 309 DMDCLMTEFGD-----IPHHAPVLAALRHLINPESTSV-- 346  
 552 DETNLFTIADBYKKFYEGLLMTVYLIEPPVPHTECKKONDLSKREHVAALIASAVCL 611  
 347 --RKIGGTALQNL-VFOYLTRLLQSLASGNDCTTSTACMCVYGLISFVLTSLELH--T 400  
 612 HYEDMNNNSREINQVLFVTFPWLKISEGN-----LMDLLKHST 653  
 401 LGNQ-QDIIITRACE-----VLAD-----PSLEPLFWGTEPPISGLGIILDSVC----- 441  
 654 LGNOVLADIESAHEELILRSILMDLRLKQLYRDLIMHAEVTAKRLAISGSGCYEYFM 713  
 442 -GMPHLLSPLL-QLLRALVSGSKTAKKY-----YSFLDKMSFYNELYKHK--PH 487  
 714 NGSSTERRRPLSDFLQIESYKVEFRVNCVLOVLDISPSTLDGGLVFLKKNAKAPN 773  
 488 DIVSHEDSTLW-RROTPKLLYP--LGGOTNLRIPOG-----TVGO-- 524  
 774 DDAYSDDSLBASSTKMGLESDFLREHSEVEIKAKRYQVQVDATHCHETSXTDGSGF 833  
 525 --VMDRAVLYRWMEYSYSWTFTCEI--EMLLHVVS-TADVIOHCORVKEPIIDLHVKY 579  
 834 INIMLTQODKLPDYDAGSVSYLNLQISVVKDLHLIGSLAVDIQVYRMHIELTDLARV 893  
 580 ISTDLSIADCLLITRSI-----YML-----LQRLTVISPPVDVIAVCNCLTVLAARPA 631  
 894 QDKNTI--CFPSVGVYIPAWYTYLISDVKKLKFVBAEVKII-----CLKV----- 938  
 632 KWTDLRHTGLFLEFAHVSYSLSQWISAEGNAGVGNLLMN--SEOPQGEVGTIAFLR 689  
 939 -----PDSSSYSPKTNGL--GYLNCFLKLEELRLSKDLIDLIDLKH 978  
 690 LITTVKCOL-----GSTOSQGLVPCVM-----FVLKEMLPYHKMRYN 728  
 979 QIESVKEGILCLRSFIDHFSSESDEHDBACGLIARVSVMAVKAEEVVIDSCIAVSHPLWYK 1038

729 SHGVEQIGCLLELILHAILNLCHET-----DLSSHSTPSLOFLCISLAATEACOT 780  
 1039 VLMISE-----VLENIKLVNVVGETCERNIYEVHVV-AKTTYVAPSPSATYGRANE 1092  
 781 VINIMIGCVDTI-DMVMAAOPRSD-----GAEGGQGGOLLIKTYKLAFTVNNVIRLKP 834  
 1093 EMEGFODTIDELKDKLQSGSPELDVIISIVGMPGLGK----- 1128  
 835 SNVVSPLQALSOQHAGHNNNLIAYLAKYIYKHDPALPRLAIQLKRLATVAPMSVVA-C 893  
 1129 -----TTLAKKTY--NDPEV-----TSREDVHAQC 1151  
 894 LGNDAAAIRDAFLTRLOSKE-----DMRKIMILEFLTVAVETOGLTEL 939  
 1152 VVTQLYSRELLITLINDVLEPSDRNEKEDBIADLRRFLITKRFLL-----LIDD 1203  
 940 FLNLEVDKG-----SDGSKFESLQWMSCLAHVALELIDSOQODRWYCPPLHRAAIAFLHA 994  
 1204 VMDYKVMNDLQMCPSDVSNRSRIILTTRLNDVAEYVKCESDP-----HH 1247  
 995 LMORRDSAMLVATKPKFMENTLSPLEGTISPSETSEP--SILETC----- 1040  
 1248 LRLFRDDES-----WTLQKEVYQGSCEPELEVDGFELEISKRCGLPLSVLVLAG 1297  
 1041 ALIMKIICLEYIYVVK-----GSLDOSLKDITLKPS-----IEKRFAYMSGYKS 1085  
 1298 VLKQKKTLIDSKKVVESLQSGRIGSLBEST--SIGTSYKNLPYVLRKPCFLYRGGFQOG 1355  
 1086 LAVHVAETEGSSCTSLLEYQMLVSAWRMLLIATTHADIMHLTDSVVRQLFLVDLGTX 1145  
 1356 KDHYSK-----MTKLM-----VAEGFVQANNEKQEDPAQGFLLDILGRN 1396  
 1146 ALLVPAVNV-----CLRSGMKCTILLI-----LQWKMRELGSV 1181  
 1397 VMMMEKRPNKVKTCRIHDLHFKCMEKAKOEPLQINSGEVFPERLEBYRLFVHSY 1456  
 1182 -DEI-----LGPLTEILEGVLOADQOLMEKTKAKVFSAFTVLQMKEMKVSIDIPOYSOL 1234  
 1457 QDEIDMRPSSNSNRSLFPMNIDPDNLMPDISIFESF-----KLVTLDLESF-- 1507  
 1235 VLVNCELTQEEV-----TALFDQTRHSIALGASATEDKDSMETD-----DCS-- 1275  
 1508 --NIGCTPTEIOVLQMKYPAQTDANSIPSIKLENTLETFFVRGLGEMILLPCSLIK 1565  
 1276 --RSRHRQDRGVCVGLHLAKELCEVDEDEDSMLQVTRRLPIPLTLT 1322  
 1566 MVKLRHIVNDRVS-FGLH-----ENMDV-LTQNSQLPNIETEST 1603

RESULT 31  
 US-09-542-331-2  
 ; Sequence 2, Application US/09542331  
 ; Patent No. 6261761  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zhong, Yi  
 ; APPLICANT: Guo, Hui-Fu  
 ; TITLE OF INVENTION: NF1 Protein and Its Role in Activation  
 ; TITLE OF INVENTION: of Adenylyl Cyclase by PACAP38-like Neuropeptides  
 ; FILE REFERENCE: 1314.1047003  
 ; CURRENT APPLICATION NUMBER: US/09/542,331  
 ; CURRENT FILING DATE: 2000-04-04  
 ; EARLIER APPLICATION NUMBER: US 09/046,745  
 ; EARLIER FILING DATE: 1998-03-24  
 ; EARLIER APPLICATION NUMBER: US 60/041,469  
 ; EARLIER FILING DATE: 1997-03-24  
 ; NUMBER OF SEQ ID NOS: 3  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 2818  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 ; US-09-542-331-2



Query Match 1.4#: Score 123.5; DB 3; Length 2818;  
 Best Local Similarity 18.5%; Pred. No. 0.26; Indels 717; Gaps 101;  
 Matches 380; Conservative 256; Mismatches 698;

QY 63 EKVANDVVASPLKELGIRISKFLGDEBOSVQLQCYIQEDYRGT-----DSVTVQ 117  
 DB 1020 EVMMARDDLSFCCEMKFR-----NMKVE-----YLDWMWGSNGQAADVDVCLTR 1066

QY 118 DERGSQLILKADVYEEERTCLACVLLITTFQDERHRYRYEVADCDKLE--KELVS 175  
 DB 1067 DLDQAS-----MEAVSLLAGL-----PLQPEBIDGVLEMAKSQLFL 1104

QY 176 KYRQOEELY-----KTEAPTWETHGNLITERQVSRFVQCLRE-----QSM 217  
 DB 1105 KYFTLFNNLINDSEVEDSAQGR--KRGSRLAS--LRICTYLANSNLINANVDSG 1160

QY 218 LLEITFL--YAYFEMASDLYITKMFKEGFGSSROTNRHLYDETMDPVDRIGYFSALI 276  
 DB 1161 LMSIGLGHYKHDQTRATEFVEVLTKIL--QOG-----TEPDILAETV--LADRFERLVELV 1212

QY 277 LVEG-----MDISLHKCA-----LDRRELIHQ-----PAQDGLQCQMD 311  
 DB 1213 IMMGOGEELPAMALANVPCSQMDLAVLYTLFESRHLTYOLMNFSSKEVELADSVQ 1272

QY 312 CL-----MLTGEDIRHNAFL--LAMALIRHTLNPEETSSVVRKIGGTAI----- 354  
 DB 1273 TLRGNSLASKIMTFCFKYGGATYLOKLDPLRLIYITSSDWQHSFEVDPRLEPSSSL 1332

QY 355 ---QLNFOYLTFLTOSLASGNDCTTS--TACMCYGLSLFVLSBELHTIG----- 402  
 DB 1333 EENQNRLLQHTETFFHAIISSEFPPLQSRVCHCIYQVQSRPONSIGAVGAMFLRF 1392

QY 403 -NOODIIDPACEVLADPSLPFLFWGTEPTSGIITLDSVCGMFPHLSPQLQTLRA----- 457  
 DB 1393 IMPAYSPYEAIGLDKKRPRIERGKLMKSK--ILQSTAN--HVLFTKEHNRPRDF 1446

QY 458 LVSGSTAKKVSFLDKMSFYNELYKRP--HDVISH-----EDGR-----LMRQTPK 504  
 DB 1447 VASNPDAAR-----FPIIDIASDCTSDAVNHSLSFISDGNVALHRLNMNGB-- 1495

QY 505 LIYPLGGQTNLRIPQGTVGQVMLDRAVILVRW-----EYSSWTLFTCEIR--M 552  
 DB 1496 ---KIGQYLSNRDRAVGRPRFDKMATLAVLGRPHKPAVDHTMSSNLNTSKFEEM 1552

QY 553 LHVASTADVIOHQRV-----KPIIDLVKAVISTDLSADCLPITSRIYML 600  
 DB 1553 TRHQVHEKEEFKALKTLISIFYOAGTSKAGNPIFYVVARFRFTQINGDIL-----LYHV 1606

QY 601 LQRLTTVISPPVDVIASCVNCLTVLAANPAKVTDLNHTG-----FLP--FVAHRY 650  
 DB 1607 LITLKPYYAKPYEIV-----VDLHTGPSNRKFTDLSKMFVVP-- 1646

QY 651 SLSLMSIABEGMNGYGNL---LMSQPOGEYGVTAFLRLITTL-----VK 696  
 DB 1647 ---GPAYNVASAVIYINCNSWVREY--TKYHERLITGLAKSKRLVFDICP 1691

QY 697 GOLGS--TOSQGLVPCVFLKEMLPSTY--KVRYNSHGVRBOIG----- 737  
 DB 1692 GKLAENHIEGQKLPALTLALEDLVKFNHMLKLAHNDQVSIKVGSTAVQVTSABERTKY 1751

QY 738 ---CLILR-----LIHAILNL----- 750  
 DB 1752 LGQSVFLNDIYVASEIEICLVNDENOPTLTIANOGTFLTFMHQCECAIVOSIHIRTRME 1811

QY 751 CHETDLSSHTPSQFLCISLAYTEAGOTVINIMGVDTIIO-----MMAAQPS 802  
 DB 1812 LSGDLSIPQHT-----KIRKQVPGTILNIALNLGSSDPSLSAAVNLCLLTCT 1862

QY 803 DGABGQGGQGLLITVLAFSVTNNV-----IRLKPPSNVVSPLQQLSQHAGN 854  
 DB 1863 FNLIKTEGO---LLETSGLCIPANNTLFIVISKTLANEPHLTLFLEBEGISFSSSIE 1919

QY 855 LIAVLAKYIYHKDPLRLAQLKRLATVAPMSVYACLGNDAA-----AIRDAFLT- 907

DB 1920 LKHLCLLEYM-----TPWLSNL-----VAFCKANDAKQORVTAIIDDKITM 1960

QY 908 ---RLQSK-----IEDMRKVMLEFLTVAVETQPLI--EFLNLEVQGS 949  
 DB 1961 TINEKQWYPSIQAKWISLGOITDL-LDVLVDSPIKTSATGGLSIRKAVWADPAVALAS 2019

QY 950 DGSKEPSLGMSCHAAVLELIDSQQDRWC---PELLHBAALAFHMLQDRDRA--M 1004  
 DB 2020 GNVLVSS--KVIGRCKIITKT-----CLSTPFLQEO-----HLMDDIATILARYM 2064

QY 1005 LVATKPKFMENTSP--LFGTLPSPSETSEPSI--LETICALIMKI-----CLEIYVYVKG 1057  
 DB 2065 LMLSFNNSLDVAAILPYLFFVVFVFLVATGPLSRASTHGLVINIHSICQSQLHF--SE 2122

QY 1058 SLDSQSLKOTLKPSIEGRFAYWS--GVYKSLANVALEBSSCTSLLEYQMLVAMRMMLI 1116  
 DB 2123 ETKQVLAISLTERBSPLEPFYLLFGLSKVKSAAV-----IA 2156

QY 1117 IATTHADIMELTDSVNRQFLDVLDO--TKALLVPAVNCIR--LGSNKCTLLILLRQW 1174  
 DB 2157 FRSSYRDRSFPSSYERETPALTSLFVTEALLEIMEBA--CMRDIPCK-----WLDQW 2208

QY 1175 -----KR--ELGSVDEILGPLETILEGVLOA-----DQO 1201  
 DB 2209 TELAQRFAPQYNSLOPRALVVFQGISKRVSHQIKQIIRILSKALESCLKQPDYNSQV 2268

QY 1202 LMEKYAKVSAEITVLOMKEMKVS DIPQYQOLVNLVNCETLQ--EYV-----IALPDOR 1254  
 DB 2269 LIBAT-----VIALTKLOPLNKDSPLHKALFWAVAVLQDDEVNLXSAGTALLQOVL 2321

QY 1255 HSLALGATEDDKSMETDCCSR--HRDQDQVCVLGLHLAKELCEVEDGDSMLQYT 1311  
 DB 2322 HTLDSIRIFENDSPEEVFAIRNPLEWHCKQND--HP----- 2356

QY 1312 RRLPILPTLLTLEVSILMKONLHF-----TEATHLILLTLARTQ 1351  
 DB 2357 -----VGLNNSNFNFALVGHLLKGYRHSRAIVANTVRIHLTLVLNKH 2402

QY 1352 QGATVAVAGITQISICLPILSYQULSTNGTAQTPSASRSKSLDAPSWPGYRLSMSLEOL 1411  
 DB 2403 RNDCKREVN--TOSVAY--LAALLTVSEVRSCLKHKKSLL-----L 2441

QY 1412 LKTLRYNPLPEALDFGVH---QERTLOCLNAVATQ--SLACLEADHTVGTILOLSNF 1466  
 DB 2442 LNDISMEVNP--MDTYPIHGEDPSYRIKAKETOPMSSPFGSGEYLAATYPTVG---QTS-- 2494

QY 1467 MKEMENHLPQLMRDIVNLGYLQACT--SLHSRMLQHYLQNKXGDLPSAVAQVQRP 1525  
 DB 2495 -----PRAKSMSLMGQPSQANTKKLGTGRKSPHLISDTK-----APKQEM 2538

QY 1526 PSAASAPSSSQPADDEASEQALHTVOYGLKLISKTLAALRHFPDVCQILLDQSL 1585  
 DB 2533 ESGITTPKMRVAVETDIEMETORISSQOHPRKRS----- 2576

QY 1586 DLAEYNPLFALSFTTPDSEVA--PSFGTLATVNVALNMLGELDKKKEPLTQAVGLST 1643  
 DB 2577 -VSESNVL-----LDEEVLTPDKIQALLTV----- 2601

QY 1644 QABGTRTLKSLMTFMENCFLYLSQAMRYLRDA--VHPRDKQMKQELSELSTLISL 1702  
 DB 2602 ---LATLVKTKITDEFDQRIIYE--YLAESAVVFPKVPVAVNHLIDSKINTLISLC 2651

QY 1703 -----SRFYRGAPSSPATGVLP-----POGSTSLSKASPSQ 1737  
 DB 2652 QDPMLNPIHGIQVQSVUYNHESPQYQOTSYLQSGRFGNLRFPAGRPBQTOI--PDVA 2707

QY 1738 EPLIQLOVQAFV 1748  
 DB 2708 ELIVKFLDALI 2718

RESULT 32

```

US-09-510-791-2
; Sequence 2, Application US/09510791
; Patent No. 6365126
; GENERAL INFORMATION:
; APPLICANT: Zhong, yi
; APPLICANT: Guo, Hai-Fu
; APPLICANT: Tong, JiaYuan
; TITLE OF INVENTION: Improvement of Learning and Short Term
; FILE REFERENCE: 1314.1047-002
; CURRENT APPLICATION NUMBER: US/09/510,791
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 09/046,745
; PRIOR FILING DATE: 1998-03-24
; PRIOR APPLICATION NUMBER: US 60/041,469
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Paeseq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2818
; TYPE: PRT
; ORGANISM: Human
US-09-510-791-2

Query Match 1.4%; Score 123.5; DB 3; Length 2818;
Beet Local Similarity 18.5%; Pred. No. 0.26;
Matches 380; Conservative 256; Mismatches 698; Indels 717; Gaps 101;

OY 63 EKVANKQVAPLKEJLGRISKFLDDEQSVQLQCYLOEDYRGR-----DSVKTVLQ 117
Db 1020 EVMAARRDLSFCQMKR-----NKWYE-----YLTBWVWGTSGQADDDYKCLTR 1066
OY 118 DERQSALLIKIADYYEERTCILRCVLLHLLTYFODEBHPYRVEYADCVDKLE--KELVS 175
Db 1067 DLDQAS-----MEAVVSLIAGL-----PLQPERGDEVLMEAKSQFL 1104
OY 176 KYRQOFSELY-----KTEAPWETHGNLNTERGVSMPVQCLRE-----QSM 217
Db 1105 KYFTLFNMILNDCSVEDESAQTGR---KQMSRRLAS--LHCTVLAMSLNLANVDSG 1160
OY 218 LLEITFL-YVAVFEWAPSDLVLTMTFKEQGFSGQTRNHLVDETMDPFVDRIGYSALI 276
Db 1161 LMSHSGLGKNDLOJTRATFMEVLTKIL-QQG-----TEFDTLAEIV--LADRFELVBLV 1212
OY 277 LVEG-----MDIESLHKCA-----LDRREELHQ-----FAODGLICQDMD 311
Db 1213 IMMGQGGELPIAMALANVPCSQMDLARVLVTLFDSHLLYQLLMMNFSEKVELADSMQ 1272
OY 312 CL-----MLTFGDIPIHNAFV--LAMLRLRHTLNPBETSSVYRKIGGTAI----- 354
Db 1273 TLFGRNSLASKIMTECFKYVGATVYLOKLDLPRLIRIVITSSQWQHSFVDDPRLPSES 1332
OY 355 ---QLNVQOYLTRLLOSLASGNDCTTS--TACMCVGLSLPVLISLEHTG----- 402
Db 1333 EBNQNLNLOHTEKFPHALIISSSEPPQLQRSVCHCLYQVQSORFPQNSIGAVGSAMFLRF 1392
OY 403 -NOODIIDTAGEVLADPSLPBLFMWTEBTSIGILILDVCGVFPHLISPLQLLRA----- 457
Db 1393 INPAIVSPYBAGILDKKPPRIERGLKJMSK---ILQSIAN---HVLFTKEHNKRPFND 1446
OY 458 LVSGSKTAKYVSFLDKMSFYNELYKHKP-HDVISH-----EDGT-----LNRQCPK 504
Db 1447 VKSNDAARR-----FFLDIASDCPTSAVNHSLSPISDGVALLHRLLMNQE-- 1495
OY 505 LLPLPGGQTNLIRPGVGOVWLLDRAVLYVR-----EYSSSWTLFTCEIE--M 552
Db 1496 ---KIGQYLSNRDKHAKVGRPPDKMALTILAVLPPEHKVPAADTWSSLNTLSKFEER 1552
OY 553 LHVSTADVLDIHCORV-----KPIIDLVHKVISTDLSIADCLPITSRIYML 600
Db 1553 TRQVHEKEEFPALQTLTISIFYQAGSKAKGNPFYVVAARFKQGINQDGL-----I 1600
OY 601 LQRLTTVISPVDVIASCVNCLTVLAARNPAKVTDLRHNTG-----FLD--FVAHPV 650

```

Db	1607	LTLTKPYAKRYEIV-----	VDLTHGPSNRFKTDFLSKMFVFR	1646	
Qy	651	SSLSQMTISAEGMNAGYGNL---	LWNSBOPOGEYGTIAFLRLITL---	VK 696	
Db	1647	-----	-GFAVDNVAAYIYVNCNSWREY--	TKYHERLLTGLKSGKRLVFTDCP 1691	
Qy	697	QOLGS--TQSGULVPCWIFUKENLPSYH--	KRYNSHGRBEOIG-----	737	
Db	1692	GKLAEBHEBEOOKLPAATLAEEDLKVFHNAKLAKDKTVKSIKVGSTAVOVTSABRTKV	1751		
Qy	738	-----	-CLIEB-----	LHAILNL-----	750
Db	1752	LGOSVFLNDIYYABEIEBICLVADENQFTLTIANQGTPELTMHQCEALVOSIHIRRME	1811		
Qy	751	CHETDLHSHHTPSLOFLCTCSLAYTEAGQVINIMIGVDTID-----	WMAAOPRS 802		
Db	1812	LSQPSJIPQFT-----	KIRPKDVBGTLTALNTALNTLGSBDSLSMAVNLCAJTC	1862	
Qy	803	DGAEGQGGQGLLITVYULAFSVTNV-----	IRLKPENNVSPLEQMLSGHAGHNN 854		
Db	1863	FNLIKBEQ--LLEBTSGLCIPANNTLTVISIKTLAANEPLTLFLEBECISGFSKSSTE	1919		
Qy	855	LIAVLAKYIYKHHPALPRLAIOQLKRLATVAPMSVYACLGNDAA-----	AIRDAFLT- 907		
Db	1920	LKHLCTEY--	-TPWLSNL-----	VAFCKNDNAKORVAILDXLITM 196C	
Qy	908	-----	-RLQSK-----	IEDMRIKWILEFLVAETOPGLI--	ELFLNLEVXGDS 949
Db	1961	TINEXWYPSIOAKIMGSLQGITDL-LDVVDSPSITKSATGGLGISKAEVADPVALMAS	2019		
Qy	950	DGSEKFSIGMMSCHAVLELIDSOQOORWY--	PLLHRAIAFLALWODRDSA--	M 1004	
Db	2020	GNVLVSS--	KVIGRMCKIKDKT-----	CLSEPTLEQ-----	HIMDDIAILARYM 2064
Qy	1005	LVLRTKRFEMENLSP-LFGLTSPSPSETSEPSI-LFETCALIMKII-----	CLSEIYYVVGK 1057		
Db	2065	LMSEPNNSLDVAALPYLFHVYTVLVTATGPISLASTHGLVINYIHSICTCSQHLF--	SE 2122		
Qy	1058	SLDOSLKDTEKKFSIEKRFAYMS--	GYVKSIAVHVAETEGSSCTSLBEOMLVSAMRLLI 1116		
Db	2123	ETKQVLRSLTEFSLPKRYLFLFGISKVSAV-----	IA 2156		
Qy	1117	IATTHADIMLTDSVVRQOLFVLDG--	TKALLVPSVNCIR-LGSMKCTLLILLRQW 1174		
Db	2157	FRSSYRDRSFGPSYERETFALTLETFEALTEIMEA--	CMRDIPTCK-----	WLDQW 2208	
Qy	1175	-----	KR-BLSGVDELGTLEILSGVLOA-----	DQO 1201	
Db	2209	TELAORFAQYNPSLOPRALVVFEGCISIRVSHGOIKOIRLSALBSCLGPPTYNSOV	2268		
Qy	1202	LMEXTKAKVFSAFITVLQKEMKMKSDIPQYSQVLVANCETLO--	BEV-----	IALFPQTR 1254	
Db	2269	LIBET-----	VALTGTQOPLANKDSILHKAALFMVAVAVLQDBVNLVYSAGTALLLEGNL 2321		
Qy	1255	HSIALGSATEDKDSMETDSCRSR--	HRDQRDGVCLVGLHAYELCEVEDEDGSMLOVT 1311		
Db	2332	HTLDSLRIFNDKSEEBEYFMAIRNPLEWHCKOMD-----	HF-----	2356	
Qy	1312	RRLPILPILTLTLEVSILMKONLHF-----	TEATLHLLTLIARQO 1351		
Db	2357	-----	VGLNFNSFNFPALVGHLLKGYRHPSPAIIVARTYRIHTLTLVNKH 2402		
Qy	1352	QGAIVAGAGITQGISICPLLSVYDLSITNGTAQTESASRKSIDAPSWGVYRLSMSLEQL	1411		
Db	2403	RNCKKFEYN--TQSVAY-LAALLVYSEVRSCSLKXKRSL-----	L 2441		
Qy	1412	LKTLRYNFLPBLDFVGVH--	QERTLOCLINAVTQO--	SIACLEADHTYGFLOLNSF 1466	
Db	2442	LTDISEMENVP--	MDPTYPIHGDPSYRTIKETQWSSPGKSGYLAATVPYV--	QTS-- 2494	
Qy	1467	MKEWHFHLPOLMRDIOVNLGYLQACT--	SILHSKRMLQHYLQONKNGGLPSAVAVORQRP 1525		

Db 2495 -----PRARKMSLDWGQPSQANTKTLGTRKSPDLISDTK-----APKRDEN 2538  
Qy 1526 PSAAASAPSSSKQPADPTEASQOALHTVQGLKILSKTALALHFTPDVCQIILDSL 1585  
Db 2539 ESGITTPPMRVARVASTYEMETORISSQOHPHLKRV----- 2576  
Qy 1586 DLAEYNFLPALFTPTFSEVA--PSFGTLATATNALNMGELDKKKEPLTQAVGLST 1643  
Db 2577 -VSESNVL-----LDEVLTPKIQALLTV----- 2601  
Qy 1644 QAEGRITKSLMTFMCNCFYLLISQAMRYLDPD-VHPRDKQRMKQELISELSTLSSSL 1702  
Db 2602 -----LATIVKYTTDEPDQRIYE---YLAASVVPKYPVAVVHNLLDSKINTLLSLC 2651  
Qy 1703 -----SRFRRGAPSSPATGVLPSS-----PQGSTLSKASPEEQ 1737  
Db 2652 QDPNLLNPIHGIVGVSVVHESPPQYQTSYLOSFGFNGLMRPAGPFSKQTOI---PDYA 2707  
Qy 1738 EPLIQVQAFV 1748  
Db 2708 ELIVKFLDALI 2718  
  
RESULT 33  
US-09-385-222A-4  
; Sequence 4, Application US/09385222A  
; Patent No. 6465211  
; GENERAL INFORMATION:  
; APPLICANT: RIKEN;  
; APPLICANT: Kanehiko Mikoshiba  
; TITLE OF INVENTION: High affinity IP3 binding polypeptide  
; FILE REFERENCE: PH-657US  
; CURRENT APPLICATION NUMBER: US/09/385,222A  
; PRIOR FILING DATE: 1999-08-26  
; PRIOR APPLICATION NUMBER: JP98/242207  
; PRIOR FILING DATE: 1998-08-27  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 2749  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-385-222A-4  
  
Query Match 1.4%; Score 123; DB 4; Length 2749;  
Best Local Similarity 18.2%; Pred. No. 0.28;  
Matches 306; Conservative 222; Mismatches 523; Indels 630; Gaps 80;  
  
Qy 2 IRKSKITSVLSFCSSRELMTILLGRSALRELSCIEELINKHMRLLGSLYKPPSPSS 61  
Db 610 ITAAEIDTFVSLVKNRE-----PRELDYLSIDLVCVSMNKS-----IPVQEL 651  
Qy 62 AEKYKANDVASPLKEGLRISKF-----LGIDEGSVOLLOQCYLOEDYRGT 108  
Db 652 ICKAVLFTNADILLETGLVLSREFEGSVSTGENALNAGEDEEB---VWLFMDSNKEI 707  
Qy 109 RD-SVKTVLQDROSQALILKTDYV-YE---ERTC-----ILRC 143  
Db 708 RSKSVRELDAQAKGQKEDRDILSYRYQNLFAWMCDROYALAINESQGLDVLRLC 767  
Qy 144 VLHLITTYQDERHRYRYEYADCC-----VDKLEKLVS-KYRQOEYELYTEA-PTWE 193  
Db 768 M-----SDENLYDYDRASFRCRLMLHMHVDRDPOEQVTPVKYARLWSEIPISEIADIDYD 820  
Qy 194 THGNLTERQVSRWVFO-----CLR-----EQSMLLEII-----FLYYAY 228  
Db 821 SSGI--SKDEIKERFAQTMERVEBYLDVQCRRFPSPDKENKLTFEVVMLANRLITFGF 878  
Qy 229 FEMAPDULVLTKNF-----KSGGFSR-QTNRHLYDETMDPFVD 267  
Db 879 YNF--SDLLRLTKILAILDCVHTTIFPISKMTKGGENKSNMNRSHIGGELMTQVVL 936  
Qy 268 RIGFSAALI-----LVGMDIESLHKALDRLR--ELHQPADQGLICQMDCLMLT 316

Db 937 RGGGFLEPTTPMAAPBEGNVKQAEBEKEDIMVMDTKITBILOFILNVRDIRISCLICI 996  
Qy 317 FGDIPHAPVLLAMALRHNTLNPEETSVVVKIIGTAIQ--LNV-----FOYLRLLOSL 369  
Db 997 F-----KREFDESNQSSETSGNSQDEGPNVPGALDPEHIEQAEIGI 1040  
Qy 370 ASGNDCTTACKCVYGLLSFVLTSLHTLGNQODITDTACEVLDPSLPFLFWGTEP 429  
Db 1041 FGGSEE-----NTPLDLDDHGR----- 1058  
Qy 430 TSGGLITLDSVCGMFPHLSPLLOL-----RATVSGK--STAKCVYS 470  
Db 1059 -TFLRVLLHTMDYDPPVLSGALQDLFRHRSQROEVLQAKQOVOLVTSQDVNDYKOIQ 1117  
Qy 471 FLDMSEFY--NELYKXK-----PHD-----VISHEDGTLW----- 498  
Db 1118 DLDDLRISIVKSELMVYKGGQPDPMDCAGSENHKKTEGTSKPLKHESTSYNRRVVK 1177  
Qy 499 -----RROTPLLYPLGGQTNLRITPQTVGOYMLDRAIYLRMEY 538  
Db 1178 EILIRLSKLCVQESASVRSRKOQORLBRMGANA-----VLE-----LLOIPY 1222  
Qy 539 SYSWTLFTCEIEMLLHVSTADVIOH-CORVAPIIDLVHKVISTDLSTADCLPITSRI 597  
Db 1223 EKADETMQ--EIKRLAH-----EFLQFCAQNOQNALHKHINFLKPGILAEAVTMQHI 1276  
Qy 598 YMLLQRLTVISPPVDVIAVCNCLTVLAARNPAKWTDLRHGFLPFVAPV---VSSL 653  
Db 1277 FMNNFOLCSFINE--RVQHFVHCIEF-HGRN-----VQYIKFLQTVKABGKFKKC 1326  
Qy 654 SOMISAEQWAG-----GYGNL--LMSBQPGS-----YGVIAFLRLITTL 694  
Db 1327 QDWMAELVNSGVDLVFNDRASFOTLIQWMSERDRMDESNPLMTHILVELLVACTE 1386  
Qy 695 VKQLGSTQSGLVF-----CWMFV-----LKEMLSYHK 724  
Db 1387 GKAVYTEIKNSLPLDDIYRVVTHEDCIPEVXIAVINFLNHCYVDTBEVMEKIYSNMH 1446  
Qy 725 WRNSHGVRQICLLIELIHAILNICHETDLHSSHTPSIQFLCTGSLAYTEAGQTVINI 784  
Db 1447 WK-----LEENFLVDICRACNNTS--DRKADS-----ILEKYVTEIYMSIVT- 1487  
Qy 785 MGIGVDITDMVMAOPRSPQAEQ--QGOGLILKTVLAFSVTNVIRLRKPSVAVPLEG 843  
Db 1488 -----TFSSPFSQOSTLLOTRQVYVQLOGVRVYH-----CWLWPSQK 1529  
Qy 844 ALSQGHAGNNLIYLAITYHKHDPALPR-LAIOU-----LKRLATVAPMSVYA 892  
Db 1530 ASVE-----SCIRVLS-----DVAKSRALAIIPVLDQVNNLFLKSHNIYQKTALMW 1576  
Qy 893 CTGNDAAIRDAFLT-----RLQ--SKIEEMRIKMWLEFLTYAVET--QPGV- 936  
Db 1577 RLSARNAARDVSLASRDYRNITIERLODIVSALED-RLRPLVQAEISLVVDYLRPELL 1635  
Qy 937 -----IEFLNLEVQDGSQKESFL 957  
Db 1636 FPENTARKKCSGGFICLKIKHTKQLEBENEKLCIKYLOTRKEMTDRGEGEQIIS 1695  
Qy 958 GMSCLHAYL---ELIDSOQODRYWCPPLHRAAIAFLHALWODRDSAMLVLRKPKF 1013  
Db 1696 D--ESENABELPOAPEAENSTEOLESPSP-----LROLEDHKGEBALRQILVNRYY 1744  
Qy 1014 W-----ENLTSPLFGTLSP--PSET--SEPSILETCALINKIICILEIYVYVKGSL 1059  
Db 1745 GNIRPSGRRSLSLTSFNGPLSPGGKPGGGGCGPSSSTSGEWSLAEVQCHLDKEGAS 1804  
Qy 1060 DOSLKDTLKKFSIEKFAVSGVYSLAVHVAETBSGCTSLLEYQMLVSARMMLLIAT 1119  
Db 1805 NLVI-DLINNASSDRFH-----ESILAIALLBEGANT-----I 1838  
Qy 1120 THADIMHLDTSVVRROLFLVDIDGTAL--LVPASVNCRLGSKCTLLILLRQWR 1176

Db 1839 QHSFECRLTEBKSEKFEKVFYDRMKVAQOEIKATVNTSIDLGNKKD-----DEVDR 1892  
Qy 1177 ELGSDVDELGLPTELEBGLADQOLMEKTA--KV----- 1211  
Db 1893 DAPSKRAKEPTQITE--EVRDQLLEASATRAFTTFRREADPDHYSOEGTOATT 1949  
Qy 1212 -----SAFTVTLQ-----MKEMKVSIDIPQSOVLVN-----VCETLQ----- 1243  
Db 1950 DKADODLEMSAVITIMQPIRLPQLCENHRDLQNPRLCQNKKNVNLVCETLQFLDCI 2009  
Qy 1244 -----EEVIALFDQTRHSLALGSAETEDKSDNETDSCSRHRDQRDQVC 1287  
Db 2010 CGSTTGGLGLGLYINERKVALINQTLSTL-----TEYCGSPCHENQN---C 2053  
Qy 1288 V 1288  
Db 2054 I 2054

RESULT 34  
US-09-228-986-78  
Sequence 78, Application US/09228986  
Patent No. 6359198  
GENERAL INFORMATION:  
APPLICANT: Strabala, Timothy  
APPLICANT: Nieuwenhuizen, Niels  
TITLE OF INVENTION: Compositions Isolated from Plant Cells  
TITLE OF INVENTION: and their Use in the Modification of Plant Cell Signalling  
FILE REFERENCE: 11000/1020  
CURRENT APPLICATION NUMBER: US/09/228,986  
CURRENT FILING DATE: 1999-01-12  
NUMBER OF SEQ ID NOS: 130  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 78  
LENGTH: 903  
TYPE: PRT  
ORGANISM: Eucalyptus grandis  
US-09-228-986-78

Query Match 1.4%; Score 122; DB 3; Length 903;  
Best Local Similarity 22.6%; Pred. No. 0.042;  
Matches 111; Conservative 66; Mismatches 168; Indels 146; Gaps 26;  
Qy 395 SLEHTLTGNO-QDIIDTAC-EVLADPSL--PELFWGTEPTSG---LGIILDSVCGM--FP 445  
Db 507 SLEVTSGNLCISFSTTACGDASSPSIEAPQVITVPERNKGNHRLAILGAVGVSLA 566  
Qy 446 HLLSPLOL-----RALVSGKS-TAKKVSFLDKMSFYNELYKHPHDYIS 491  
Db 567 ILILPLFVFMYRRRGRTSMYTERAVADVNRWMAKIFSYKEIKATNNF-----KEYIG 621  
Qy 492 HED-GTLWRROTPKLLYPLGGQTNLRIPQTVGOVMDRAVLVWEXYSWTLFTCEI 550  
Db 622 HGSFSGVYLGMP-----VGKLV-----AVKRAFDTQLGADSFINEV 659  
Qy 551 EMLLHV-----VSTADVIQHCORVKEPIIDLVKVISTDLSIADCLPITSRIYML--LQR 603  
Db 660 RLISQVRHQNIVLSLEGFCYESQRQ---ILVEYELPG-SLADQLYGNRSKRFSLSWVR 714  
Qy 604 LTTVISPVDVYIASCUNCILTVLAARNPAKWTDLRHTGFLPFVAHPVSSLQMSIAEGMN 663  
Db 715 LK-----IADDAKG-----LDVLIHNSNPRRIIHRDIKCSNIILDKEMN 753  
Qy 664 AG--GYGNLMMNSBOPOGEYGVTTAFLRLITTLVKGQSGTOSQGLVPCWVFLKEMLP 721  
Db 754 ARLCDFG--LSKOMIQDPATH-----VTVVAGTAGYLD-----PE 787  
Qy 722 YHKRRYNSHGVREQ-----IGCLILELHAILNLCHETDLSHSTPSIQFLCISLAYTE 776  
Db 788 Y-----YSTQQLTEKSDVYSFGVVLLELI-----CGREPLNHSGRPDSFNLVLMKAPYLQ 837  
Qy 777 AGQTVINIMIGVDTIDVWMAAOPRSDBAGGQGLIKTVKLAFSVTNNVIRLKP--S 835  
777 AGQTVINIMIGVDTIDVWMAAOPRSDBAGGQGLIKTVKLAFSVTNNVIRLKP--S 835

Db 838 AGAFEIVDESIG-----GSFVDSMRKVAKIAVRSEVERDASIRPTIA 879  
Qy 836 NVSPLEQALS 846  
Db 880 QILSVLKEAYS 890

RESULT 35  
US-10-101-464A-78  
Sequence 78, Application US/10101464A  
Patent No. 6768041  
GENERAL INFORMATION:  
APPLICANT: Strabala, Timothy  
APPLICANT: Nieuwenhuizen, Nicolaas  
APPLICANT: Higgins, Colleen M.  
TITLE OF INVENTION: Compositions Isolated from Plant Cells  
TITLE OF INVENTION: and their Use in the Modification of Plant Cell Signalling  
FILE REFERENCE: 11000.1020C2  
CURRENT APPLICATION NUMBER: US/10/101,464A  
CURRENT FILING DATE: 2002-03-18  
PRIOR APPLICATION NUMBER: 09/704,302  
PRIOR FILING DATE: 2000-11-01  
PRIOR APPLICATION NUMBER: 09/228,986  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/162,866  
PRIOR FILING DATE: 1999-11-01  
PRIOR APPLICATION NUMBER: PCT/US00/00724  
PRIOR FILING DATE: 2000-01-11  
NUMBER OF SEQ ID NOS: 989  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 78  
LENGTH: 903  
TYPE: PRT  
ORGANISM: Eucalyptus grandis  
US-10-101-464A-78

Query Match 1.4%; Score 122; DB 4; Length 903;  
Best Local Similarity 22.6%; Pred. No. 0.042;  
Matches 111; Conservative 66; Mismatches 168; Indels 146; Gaps 26;  
Qy 395 SLEHTLTGNO-QDIIDTAC-EVLADPSL--PELFWGTEPTSG---LGIILDSVCGM--FP 445  
Db 507 SLEVTSGNLCISFSTTACGDASSPSIEAPQVITVPERNKGNHRLAILGAVGVSLA 566  
Qy 446 HLLSPLOL-----RALVSGKS-TAKKVSFLDKMSFYNELYKHPHDYIS 491  
Db 567 ILILPLFVFMYRRRGRTSMYTERAVADVNRWMAKIFSYKEIKATNNF-----KEYIG 621  
Qy 492 HED-GTLWRROTPKLLYPLGGQTNLRIPQTVGOVMDRAVLVWEXYSWTLFTCEI 550  
Db 622 HGSFSGVYLGMP-----VGKLV-----AVKRAFDTQLGADSFINEV 659  
Qy 551 EMLLHV-----VSTADVIQHCORVKEPIIDLVKVISTDLSIADCLPITSRIYML--LQR 603  
Db 660 RLISQVRHQNIVLSLEGFCYESQRQ---ILVEYELPG-SLADQLYGNRSKRFSLSWVR 714  
Qy 604 LTTVISPVDVYIASCUNCILTVLAARNPAKWTDLRHTGFLPFVAHPVSSLQMSIAEGMN 663  
Db 715 LK-----IADDAKG-----LDVLIHNSNPRRIIHRDIKCSNIILDKEMN 753  
Qy 664 AG--GYGNLMMNSBOPOGEYGVTTAFLRLITTLVKGQSGTOSQGLVPCWVFLKEMLP 721  
Db 754 ARLCDFG--LSKOMIQDPATH-----VTVVAGTAGYLD-----PE 787  
Qy 722 YHKRRYNSHGVREQ-----IGCLILELHAILNLCHETDLSHSTPSIQFLCISLAYTE 776  
Db 788 Y-----YSTQQLTEKSDVYSFGVVLLELI-----CGREPLNHSGRPDSFNLVLMKAPYLQ 837  
Qy 777 AGQTVINIMIGVDTIDVWMAAOPRSDBAGGQGLIKTVKLAFSVTNNVIRLKP--S 835  
Db 838 AGAFEIVDESIG-----GSFVDSMRKVAKIAVRSEVERDASIRPTIA 879  
Qy 836 NVSPLEQALS 846

```

Db      880 QILSVLKEAYS 890

RESULT 36
US-10-028-946-4
; Sequence 4, Application US/10028946
; Patent No. 6734009
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; APPLICANT: Miranda, Maricar
; TITLE OF INVENTION: Friddle, Carl Johan
; FILE REFERENCE: LEX-0289-USA
; CURRENT APPLICATION NUMBER: US/10/028,946
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/258,335
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1958
; TYPE: PRF
; ORGANISM: homo sapiens
US-10-028-946-4

Query Match      1.4%; Score 122; DB 4; Length 1958;
Best Local Similarity 18.1%; Pred. No. 0.18;
Matches 274; Conservative 226; Mismatches 544; Indels 468; Gaps 67;

Qy      441 CGMPHLISPLLIQILRALVSG-----KSTAKKVSFLDKMSFYNELYHKPHDVI SHEDG 495
Db      105 CGHFEE-----VQVVRKATGDIYAMKMKKALLAQGVSEFFEE-----ERNILSRSTS 154

Qy      496 TLMRQTEKLYPLPGCGQINLR-----PGTVGQVMDLDRAYLVIRMEYSYSMTL--FTC 548
Db      155 P-W--ITQLQAFODKNHLYLVMEYQPG-----DLGLNLYEDQDLNLIQFYLA 203

Qy      549 EIEMLLVHVSADVLIQHCRVPE--IIDLVHKVISTDLIA-----DCLLPITSR 596
Db      204 ELILAVSHVLMGVYH--RDIKPNILVDRGHILVDFGSAKNSNMKNVNAFLPICTP 261

Qy      597 IYMLIQRLITTVISP-----PVDVASCNCLTVLAARNPAKVTWDLRHHTGFLP 644
Db      262 DYMAPEVLTIVNAGDKGTGGDWMWSGVIA--YEMTIGRSP----- 302

Qy      645 FVAHPVSLSQMISAEKNGAGYGNL-----LMSSEQGEYGTIAFLILITTVYQG 698
Db      303 -----FABGTSAKTFNNININFORFLKFPDDPK-----VSDFLDLISLQCG 345

Qy      699 LGSIQSQGLVPCVMEFVLKEMLPYHKKRYSNGVAREQICLILIELIHAILMICHETDLHS 758
Db      346 KERLKEEGL--CCH-----PFFSKIDMN--IRNSPPF-----VPLKSDDDTSN 387

Qy      759 SHTSLSQLCLCSLAATYAGQTVINIMIGVDTIDMWAAPRSDGAEQCGCG----- 811
Db      388 FDEBEKN-----SWSSSSPCQLSPSGFSGEBELPFGVFSKALGILGRSESVSGIDS 440

Qy      812 -----QLLITVKLAFS-----VTNNVRLKPPSNV 838
Db      441 PAKTSMKELLISKELQDSODCKHMEQEMTLHRVSEVAVLISQKVELAASQTOR 500

Qy      839 SPLSQALSQHGAGNNLIALV--AKYIYKHDPALPRILAIQLRLATVAPMSVYACIGN 896
Db      501 SILRQDLATYITECSLSKRLSEQARMEVSGEDDK--ALQLLDIR-----E 544

Qy      897 DAALIRDAFLRLRSKIEDMRIXMILIEFLVVAETQGLIELFLANVEDGDSQSEPS 956
Db      545 QSRLOETKEQEQAYQVEEMRLMNQLEEDIVSARRRDLIE--SELRESRLAAEFK 600

Qy      957 LGMWSCLH-----AVLELDSQOQDRYM--CPPLIRHRAALFLHA--LMQ 997
Db      601 RKATECQHKLKAKDQKPEVGEYAKLEKINAEQOLKIQELQETLERKAVASTATELLQ 660

```

```

Qy      998 DRDPSMLVLRTRKPFWENLSPFLGTLSPSESEPSILET-----CALMKITCLEIYY 1053
Db      661 NTRQKERAERELEKLNREDS-----SEGIRKKLVAEERRSHLENKVRLE-- 708

Qy      1054 VVKGSILDSKDTLKKEFSIEKRFAYMSGYSLAVHAETGSSCTLLLEYQMLVSAMRM 1113
Db      709 -TMRERNRLKDDIQTSQCG-----IQVADKLELEBKH-----REMQ 746

Qy      1114 LLIATTHADIMHITDSVVRQLFDVLDGFKALLLVPAVNCILRQSMKCTLLILIRQ 1173
Db      747 ---VSAQHLLEV-HIKQKQEHYEERKIKVLD-----NQ 773

Qy      1174 MKRELGSVDDELGLPLTILLEGVQADQOLMEKTRAKYFS--AFITVLOMK-----EMKVS 1226
Db      774 IKQDLAD-----KETTENNMQRHEEZANH--KGKILSRQKATINANDSKINSLEQRIV 824

Qy      1227 DIPQSQVLNVG-----ETLQEVYALPDRQHSIALGASATEDKDSMETDD--CSRSRH 1279
Db      825 ELSEANKLANSSILFTQRNKAQEBMISELRQGFLETQAGKLEAQRKLEBLEKISH 884

Qy      1280 RDQRDGYCVGLHIAKELCEVD--EDGDSWQVTRRLPILPTLITLLEVSILRMKN---L 1334
Db      885 QDHSQKRL--LELETRELVSEHEBQKLELRQ-----LTELQSLQERESQUTAL 935

Qy      1335 HFTFATLHLITLARTQAGATAVAGAGITQISICPLSYLVQYSTNGPAQTSPASRSKSLDA 1394
Db      936 QAAAPALLESQRLQAKTELETAEAEETIAL-----TANHDEIQRK--FDA 980

Qy      1395 PSMPGVYRLSMS-----LMEQLLTLRYN-----FLPBAIDFV--GVHQ----- 1431
Db      981 -----LRNSCTVYTDLEBQNLQTEDNAELNQNFFYLSKQDLBAGANDEIVQLRSEVD 1034

Qy      1432 -----ERTLOCINAVRQVSL--AC-----LEBADHTVGEIILQSLN 1465
Db      1035 HLRREITERMQULTSQKQIMEALKTTCTMLEBOYMDIEALNDELLEKEREQWEAMRSVLGD 1094

Qy      1466 FMKEWHFHLPOLMRDIQVNLGYLQCACTSLHSRKMLOHYLQNNCGD--GLPSAVARQV 1523
Db      1095 EKSGFEVRVRELQMDTEKQSRARADQRTESRQVVELAVKHKKIILALQALKEQKL 1154

Qy      1524 RPPSASAPSSSKQRA--ADTASEQALHTVQYGLKILSKTLAALRHFTPDVCQIIL 1581
Db      1155 KASISDKLNDLEKKAHLEMNARSLOQKLETERELKORLIEE-----QAKL 1201

Qy      1582 DQSLDLAEVNFALSTPTTPFSEVAPSRGITLATVNNVALMGLDKKEPILQAVGL 1641
Db      1202 QQQQNDL-QKNHITRLT-----QGLQEBALDRADLDLKTERSDL 1236

Qy      1642 STQAEGRRLTKSLIMFTMENCFYLLISQAMR--YLADPAVHPDRKOR----- 1686
Db      1237 EYQLENIQVLYSHEKYMGB-----TISQOTKLDFQAKMDQPAKKKVLYQNELKAL 1292

Qy      1687 -----MKQELSELSLTLSSLSRYRRGA-----PSSPVT-----GVLPSPQG 1724
Db      1293 EKERKARCAELAELOKTRIELRSAREBAARAKATDHPSTPATARQIAMSALVRSPEH 1352

Qy      1725 KSTSLSKASPS 1736
Db      1353 QPSAMSLAPPS 1364

RESULT 37
US-10-028-946-2
; Sequence 2, Application US/10028946
; Patent No. 6734009
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; APPLICANT: Miranda, Maricar
; TITLE OF INVENTION: Friddle, Carl Johan
; FILE REFERENCE: LEX-0289-USA
; CURRENT APPLICATION NUMBER: US/10/028,946-

```

```

; CURRENT FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: US 60/258,335
; PRIOR FILING DATE: 2000-12-27
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2054
; TYPE: PRT
; ORGANISM: homo sapiens
; US-10-028-946-2

Query Match      1.4%; Score 122; DB 4; Length 2054;
Beet Local Similarity 18.1%; Pred. No. 0.2;
Matches 274; Conservative 226; Mismatches 544; Indels 468; Gaps 67;

QY 441 CGMPHLLSPILLQILRALVSG-----KSTAKKVSFLDKMSFYNELYHKPHDIVSHEDG 495
DB 105 CGHFAE-----VQVVRKATGDYAMKVMKKKALLAQGVSFEE-----ERNILSRSTS 154
QY 496 TLMRQTPKLYPLGGQTNLRI-----PGTVGVMLDDRAVLVWEXSYSSWTL--FTC 548
DB 155 P-W---IPQLQYAFODKNHLYLVMEYOPG-----DLTLNKKYEDQDLNLIQFYLA 203
QY 549 EIEMLHVSADYVIOHCORVYP---IIDLVHKVISTDLIA-----DCLLPITSR 596
DB 204 ELILAVSHVHLMGVYH--RDIKPENIIVDRGHITLVDPFGSAAKNSKMWNAKLPICTP 261
QY 597 IYMLQRLTVVISP-----PVDVASCVNCITVLAARNPAKVTDLRHHTGFLP 644
DB 262 DYMAPEVLTVMNGDGKGTGGLDCDWMVSGVIA---YEMTYGRSP----- 302
QY 645 FVAHPVSLSQMISAEKNAGGYGNL-----LMSBPQGEYGTIAFLILITTVYQG 698
DB 303 -----FAEGTSARTENNINMFORFLKFPDPK---VSDFLDILQSLGCG 345
QY 699 LGSTOSGLVPCVMEFVLKEMLPYHAKMRYNSHGVEQIGCLILIHAILNI,CHETDLHS 758
DB 346 KERLKFEBL--CCH-----PFEKIDMNN--INSPPPP-----VPLKSDDDTSN 387
QY 759 SHTESLOFLCISLAYTAGOTVINIMIGIVDTIDMVAQPRSDGAGGCG----- 811
DB 388 FDEBEKN-----SWSSSPQQLSPSGFSGBELPFGVFSYKALGIGRSESVSGDS 440
QY 812 -----QILITVKLAFS-----VTMNVIRLKPSPNV 838
DB 441 PAKTSSMEKLLIKSKELQDSQDKCHKMOEWTLRHREVEAVLSCKEVELKASETOR 500
QY 839 SPLRQALSQGHAGNNLIALV--AKYIYKHDPALPLAIDLLKRLATVAPMSVYACIGN 896
DB 501 SLLEDLATYTTTECSSLKRSLEQAMEVSQEDDK---ALQDLHDIR-----E 544
QY 897 DAARIRDAFLTRLOSIEDMRIKWILFELTVAVETOPGLIELFLMEVKDSDSKES 956
DB 545 QSRKQEIKEQEQYQAQVEEMRLMWNQLEBDLVASARRSDVY---SLRBSRLAAEFK 600
QY 957 LGMNSCLH-----AVLELDSQODRYM--CPPLHRAAIAFLH--LWQ 997
DB 601 RKATCCQHKLKAKDQGRVEGEYKLEKINABOOLKQOELOEKLEKXVKASTETELLQ 660
QY 998 DRDRSAMVLRTKPFEMNTLSPGLTSPSETSEPSILET---CALIMKIICLEY 1053
DB 661 NIRQAKERAERLEKQNRDS-----SEGIRKQLVAREERHLEKVKRLE--- 708
QY 1054 VVKSGLDQSLKDTLKKSIEKRFAYMSGYVNSLAVVAETGSSCTSLLEYQMLVAMRM 1113
DB 709 -TMRERENRLKDDIOTKSOQ-----IQOMADKILELEBKH---REAQ----- 746
QY 1114 LLIATTTADIMHLTDSVVRQLFDVVDGTAKALLVAVSNCLRLSGMKCTILLILRLQ 1173
DB 747 ---VSAQHLLEV-HLKQEQHYEEKIKVID-----NQ 773
QY 1174 WKREIGSVDELIGPLTEILEGVLQADQOLMEKTKAVFS---AFITVLQMK---EMKVS 1226

```

```

DB 774 IKKDLAD-----KETLEMMQHRHEBAHE-KGILISEQKAMINMDSKIRSLREQIV 824
QY 1227 DIPQYSQVLAVNC-----ETLQSEVIALPQTHSLALSGATDKDSMETDD--CNSRH 1279
DB 825 ELSBANKLANNSLPTORNNKKAQOEMISELNQOQFYLETQGLKLEADRKLKEOLEKISH 884
QY 1280 RDQRDGVCVGLHHLAKELCEVD--EDGSDWLOVTRRLPIPLTLTTEVLSMKON----L 1334
DB 885 QDHSDDKRL--LELETREVSLEHEOEKLELKQ-----LELEQSLQDRSSQLTAL 935
QY 1335 HPTBATLHLTLTARQOGATVAVAGITQISICPLSLVYQSTNGTAQTPSASRKSIDA 1394
DB 936 QAAABALESQURQAKTELEETTAEBEEIQTL-----TAHRDEIQRK--FDA 980
QY 1395 PSWPGVTRLSMS---LMEQLKTLRYN-----FLPEALDPV--GVHQ----- 1431
DB 981 -----LNSCTVITDLEBQNLQJTEDNVAELNQNIFYLSQDLBASGANDIYOLRSEVD 1034
QY 1432 -----ERTLQCNNAVTVQSL--AC-----LEADHTVGFILQLSN 1465
DB 1035 HLRREITEREMQLTQKOTMEALKTTCMLBEOVMDEALNDELLEKROHEAMRSVLGD 1094
QY 1466 FMKEWHFPLPOLMEDIQVNLGYLQACTSLHSRKMLOHYLQNRQGD--GLPSAVARVQ 1523
DB 1095 EKSQFECGRVRELQWMLDTERKQSRABDQRTESQVVELAVKEHKAETIALQOLKEQKL 1154
QY 1524 RPPSAAASAPESSSQPA--ADTEASBOQALHTVOYGLIKILSKTIALRHFTPVQCLL 1581
DB 1155 KAESISDKLNDLEKHMALKNARSLQOKLETREBELKORLEB-----QAKL 1201
QY 1582 DQSLDLAEVNFPLFSFTTPTFDESEVAPSFGLTATVAVNALMGLDKKEPTLOAVGL 1641
DB 1202 QOQMDL--QKHIFKLT-----OGLEALDRADLLATERBDL 1236
QY 1642 STQAEGRTRLKSLMFTMENCFYLLISQAMR--YLBDPAVHPDKOR----- 1686
DB 1237 EYQLENIQVLYSHKVMKMG---TISQOTLIDFLQAKMDQPAKKKVPLOQVNEKLAL 1292
QY 1687 -----MKQELSESLTSLSSLSRYFRGA-----PSPPAT-----GVLPSQOG 1724
DB 1293 EKERARCALEBEALQKTRIELRSARBEAMARKATDHPHPTPATARQOIAMSAIVRSPH 1352
QY 1725 KSTSLSKAPES 1736
DB 1353 QPSAMSLAPPS 1364

RESULT 38
US-08-952-127-12
; Sequence 12, Application US/08952127
; Patent No. 621136
; GENERAL INFORMATION:
; APPLICANT: Shilo, Yosef
; APPLICANT: Tagle, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: ATRAXIA-TELANGIECTASIA GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 621136thwestern Hwy., Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/952,127
; FILING DATE:
; CLASSIFICATION: 530

```



```

ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,995
REFERENCE/DOCKET NUMBER: 2290.00029
TELECOMMUNICATION INFORMATION:
TELEPHONE: 810-539-5050
TELEFAX: 810-539-5055
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3066 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Mus musculus
US-08-952-127-12

Query Match 1.4%; Score 122; DB 3; Length 3066;
Best Local Similarity 17.8%; Pred. No. 0.43;
Matches 298; Conservative 234; Mismatches 552; Indels 592; Gaps 81;

QY 89 DEEGSV--QLQCVYQEDYGRGTRDSVKTIVLQDERGSQLIKIADYYEERTCILRCVH 146
DB 1269 DEWKSANOIQKCM-----KSLVD-----CEPKLIH 1296
QY 147 LLYFQDE--RHYRYEYADCVKLEKELVSKYRQGFELYKTEAPTWETHGNIMTEROV 204
DB 1297 ILPAYAYGTDYS-----VQKRETKATKYDT-----LKGEDFLKQI 1335
QY 205 SRMFQCCRBOGSMLEITFLYYAYFEMAPSDLLVITKFKEGRGSRQTNHLYDET--M 262
DB 1336 DQVFISNLP--IVVELMTLHETADSADSD-----ASQATALCDFSGDL 1379
QY 263 DPFVDRIGYFSAFLIVEGMIDIESLHKCALDDRBELHQFADGILIQDMDCMLTFGDIR 321
DB 1380 DPAPRPFPSPHVIOATFAYISNCKTKFKSILEI-----LSKIPD 1420
QY 322 HHAPVLLMALLRHTLNPEETSSVVRKIGTAIQLVNQ--YILRLQSLASGNDCTTST 380
DB 1421 SYOKILLIIC-----EQAAETNNVPK--HRIKIYHLPVSLIKDIOGSGGAMAFV 1471
QY 381 AOCMCYGLSFV-----LTSLEHTLGNODIIDTACEVLADSLPELFGTEPTGGLG 434
DB 1472 LRDIYITLHYINKRSSHFTDVSLSFSLCCDLLSRVCHTAVTCKAL-----ESH 1525
QY 435 IILDVSCGM--FPHLSPLQLLRALVSGKS-----TAKKVSFLDKMSFYNEL----- 481
DB 1526 VIVGTLPLVDYQYQEQGVLDLKLIVIDKNKNLSTITKLDLPFDHVF--KDLRLQ 1584
QY 482 ----YKAKHDVISHEDGT--WRQTPKLYPLGQQTNLRIPGTGVQVMLDDRAYLVW 536
DB 1585 QKIRYSGGPFSLIEINHFLSVSAVNPPLRLTLEGLKDLRQLRQHKDQMD--LLRA 1640
QY 537 EYSYSSWTLFTCEMLHVVSTA-----DVIGHC--QRVKPI-----IDLVHVIS 581
DB 1641 SQDPQDSIVVKLVLSLQTSKMAVNOTGEREVLEAVGRCLGELGIDPSTIAQR--N 1697
QY 582 TDLISADCL--LPTSRIYMLQRLT--TVISPPVDIASCVNCL--TVLAARNPAKWT 635
DB 1698 KDVSYTKAYGLPDRRELQWTLIMLTALNNTLVEDSVKIRSAACLNILATIKGHFWE 1757
QY 636 DLK-----HTGLPLFVAHPYSSLSQMTISABGNAGGYGNLANSQPGQE-- 680
DB 1758 NYKTSADPMTLYLOPFTSRKKPFLVPSVKEVLT--EGDLA-----VNLWVPGESH 1808
QY 681 ----YGVTLAFRLITLTVKQGLSTQSGVPCVMFVLEKMLPSYHKMYNSHGVAEQI 736
DB 1809 DIWKITITCAPL-----DSGINSITLQILKPMCE-----YATDF 1843
QY 737 GCLILE-LIH-AIINLCHET--DIHSHSTSLQFLCISLAYTEAGQTVININGIGVDIT 792
DB 1844 COMLLPYLIHDVLLQDTHESWRTLSSAVRG--FTSCFPGHSSQASR----- 1888

```

```

QY 793 DMWAAQPRSDGAGGOGGOLL--KTVKLAFTVNNVILKPPSN----- 836
DB 1889 ----SATPANSDESENFELRCCLDKKSQRTMLAVDYLRRQKPPSGTADPAWFLDINY 1944
QY 837 -VVSPLRQALSQH-----GAHGNLLIYAKYIYKH 867
DB 1945 LEVAKVAGSCSAHNTALLVAEISDKSTDBQEKSPSTPEEGSGTT--ISLSKSEET 2003
QY 868 DPALPRLAIOULKRLATVAPMSYACIGNDAAIRDAFLTRLOS--KIDNRKIMILEFL 926
DB 2004 GISLQDLLEIYISIG--EPDSLYGCGGGMLOP-----LTRIRYHEATWEXALVYDL 2057
QY 927 TVAVET---QGLIEFLPNEVVDGSDGSKSEPLGMMSCAHVLELDSQOQRYWCPPL 983
DB 2058 ETSISSSTROSGIIOALQN-----LGSHIISLYVKGLDYER--REWCAL 2101
QY 984 LH-RAAIAFLHALW-----QDRDSAMLVLRTK--PKFMENTSPLEFGL 1025
DB 2102 QELRYQAAHNMQWGLCASAGQVEGTSYHESLYNMLQCLRNREFSTFYESLR---YASL 2158
QY 1026 SPSPSESEPSILTCALIMKILCLEIYVYVKSGLDQ--SIKDTLKRFST----- 1072
DB 2159 FRVKEVEELS-----KGSLESYSLYPTLSRLQALGELENGELF 2198
QY 1073 ----EKRFAYW-----SGVKSIAVHVAETGSGSCTS-- 1100
DB 2199 SRSTVDREREBAWKKOKKSQSLKQDPSFOEPLMLRTYILELVOKEMERGQACSKD 2258
QY 1101 ----LLEYOMLVSAW-----RMLLIATTHADI-----MHLDTSV-----RRQLFL 1138
DB 2259 ILTKHIVEFSLARTFENQPLPERAIPKIQVNSALCGISEMHLBEAQVWAKQESLAL 2318
QY 1139 DVDL-----GFKALLVPASVNCIRL--GSM-----KCTLLILLR 1172
DB 2319 SILKQMIKKIDSSFQKENDAGIKVIV-----AECLRVGSGWLAETCLENPAVIMQTYLE 2373
QY 1173 QMRRELSYDEILGPLEILLEGVLQADQOLMEKTKAKVSAFTIVQOMKEMKVSIDIPOY- 1231
DB 2374 KAVKAVAGSYD--GNSRELBNQMK-----AFUSLARFPDQVQRIENY 2415
QY 1232 ----SOLVNVGCTTLOBEVIALFDQTRHSLALASAT--EDKDSMETDCCSRHRDOR 1283
DB 2416 KSESEFNKQYTL--KRAKEEVGLL--REHKIQTRVYVVKQRELEDECARLALREDR 2469
QY 1284 DGVCVLGHLAKELGCV-----DEGDSWLOVTRRLPLPLTLTLEVSLEMKON 1333
DB 2470 ----KRLCAVENYINCLISGEHDLW--VFRLCISWLENSGVSEVNGMKKO 2517
QY 1334 LHFTATLHLTLTARTQOGATAVAGAGITQSGICPLSVYOLSTN--GTAQTPSASRSL 1392
DB 2518 ----GKISYKFLPLM--YDLAAMGTRKMG----- 2544
QY 1393 DAPSWGVRRLMSLMEQLKTLRYNLPBALDPVGVQHOERTLOCLNANVTQSLACLEE 1452
DB 2545 ----LGFHB-----VNNILISRLSH 2562
QY 1453 ADHTVGFILQSNFMKEMHFLPOLMRDIOVNLGY-----LCOACTSLHS 1498
DB 2563 PHHTLFIILANANKDEFLSKPRTTRSRITKSTSKENSHLDEDRTEATRIIHS 2618

RESULT 39
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores

```



```

STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,254
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/141,239
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 1191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-9949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2482 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-328-254-6

```

Query Match 1.3%; Score 121; DB 1; Length 2482;  
 Best Local Similarity 18.5%; Pred. No. 0.36;  
 Matches 395; Conservative 329; Mismatches 743; Indels 668; Gaps 105;

18 RELVTLLGRSALRELSQIEALNKGWRRLLEGLSYKPPSPSAEKYKANKDVAPLKE 77  
 48 RLNLNVLDSK-----VEVETOK-----LAVMELQOKAKESDDQKQKEI-----EN 88  
 78 LGLRISKPLGL-DEEOSVOL-----OCY--LQEDYRGTRDSVK-----TVLQDERQ 121  
 89 MCLKTSQUTGVDELEHKLQLLSNEIMDKRCYQDLHAYESLRLKSKXASLVTNEDH 148  
 122 SQALILKADYYEERTCILKCVLHL--TYFQDERPRPYE-----YADCYDK 168  
 149 QRSILA-----FDQOPMHHSFANIIGEOSMPSESECRLEADQSPKSAIILQNRVDS 202  
 169 LE-----KELVSKYKROQFEELYKTEAPTWETHGNLMTEROVSWMFY-----QCLRE- 214  
 203 LEFSLESGKQMSDYLQKCEELVQIKG--EIEENMLKAEQMHQSFWAETSORISKLOED 259  
 215 -----QSMLEIIFLYAAYFEMAPSDLVLT--KMFKEQG-FGSRQTNRLVDETMDPFYDR 268  
 260 TSAQNVAAYELT---SALENKEKELQLNDKVEFEQAIEQLKKSNNILDESU---KE 311  
 269 IGYSAIILVGEAMDIESLHKCALDRRELHOPADQGLICODMOCMLTFGDIPIHAPVLL 328  
 312 LQLSETSLERKEKMSIISL--NKREIEBELTQNGTLKEINA----- 352  
 329 AMALLRHLNPEETS SVVRKIGTALQNL-----VFOYLTR--LLOSLSAGN-- 374  
 353 -----SLN-QEKNMLIQKSESFPANYIDEREKSISELSDOYKQKLLILQRCSETGNAY 404  
 375 -----DCTTSTACM--CYUGLSFVLTSLELHTLGNQODIIDTACEVLA--D 417  
 405 EDLSQKYAAQAKNSKLECLNLC-----TSL-----CENRKNELQKLEAKAKHQ 451  
 418 PSLPELFGTEPTSGLIILDSVCGMPHLLSPILLQILRALY-----SGKSTA----- 465  
 452 EFLTGLAAEERNQMLLELFTV-----QALRSEWTDNQNNKSKSAGAGKQEI 500  
 466 -----KKVVSFLDKMSFYNELYKHKPHVDYISHBDGLYMRQRPKLLLYPLGGQTN 514

```

501 MTLKEQNKKQKENVNDLLQENBQLMKVMKTK-HECQMLESEPI--RNSVKERESEBRQCN 557
515 LRIPQG--TYGQVWLD-RAYLVMEYSYSWTL-----PTCEIMLHVST--ADV--- 562
558 FK-PQMDLEVAWEISLDYVNAQLVQLEAMLRKKEIKQSEKEXKCLHELTQINGDLETS 616
563 -----IOHCORVKEIIDLVHKVTS-----TDLSIADCLLEPITSRIYML 601
617 NLQMOGQOEISGLQDCE-----IDAEKYISGFHELTSONDANHLOCSLOTTHAKNEL 671
602 QRLTTVISPPVDVYASCVN-----CLTVL--AARNPAKWTDLRHTGFLPFVAHPVSSL 654
672 EKICEIILQAEKYELVTELNDRSRSECTATRMGAEEVGKLLNEVK-----ILNDSGGL 724
655 QMISAEQNMAGCYNNLMSNEQPOGEYGVTTA-----FLRLITLVKQGLSGTSQ 705
725 HGEIVEDI PGGEFG-----EQPNEOHFVSLAPDESNSEHLLTSDKEVQMHFAELQEK 778
706 GLVPCWFWLKEMLPSYHKMRVNSH-----GVREQIGCL 739
779 FL-----SLQSEHKLIHDQCOMSSMSELQTYVDSLKXENLVSLNLRNFOGDL 828
740 ILLELIHAILNICHETDLSHSHTPELQFLCI-----CSLAYTEAGQTVINIMIGVDTID 793
829 VKEM-----QLGIEBGLVPSLSSCVDDSSLSLSDGSSPYRLLEQTG-----D 873
794 WMAAQRPSDGAEGOGQQLIKTVKLAFSVTNNVIRLKPSPNVVSPLEQALSGHAGN 853
874 MSLLS--NLBGA--VSANQCSVDDEVFCSLSQIEENLTREKTPSAPAKGEELSECEVVRQ 929
854 NLIAVIAKYIYHKHIDPLRLAIOLEKRLATVAPMSVYACIGNAAIIRDAFLRLQSKI 913
930 SLEKLEEM--ESQGITRKKEIOELQDLS-SEROELDCRKQYLSNEBMOQKLSVT 985
914 EDWKIKWILEFLT--VAVETOPGLIEP-----FLNLEVXDGSGSKFSLGMMWSC 962
986 LEMSKLAABKKQFQESLSEIVARLQGLDLSRSLIGITDTEADLAGRNE-----SC 1039
963 -----LNAVLELID-----SQQDRYW 979
1040 DISKHESETTERPRPKHDVHQICDKDAQDNLMDIEKITEGALKPTGECGSGEOSPDTNY 1099
980 CRP-----LHRAATLFLHAL-----WODRDSAMVLRTKPKFMENL----- 1017
1100 EPPGEDKTQSSBECISLSESGFNALVPMDFLGQEDIHNLQLVKETSNNELILHYIE 1159
1018 --TSPFLGTLSPPESETSEPSILETCALIMKI-ICLEIYVVYVGSLL--DOSLKDTLKXS 1071
1160 DDRKRVSLNEMKELEDSKHLQLEVOQMTKEIACIELEKIV-GEKKENSPLSEKLEVFS 1218
1072 IE-----KRFAYWGSYKSLAVH-----VAETBGSSCTSLLEYOMLVSANRML 1114
1219 CDHOBELQRVVETSGLSNDLEMHADKSREDIGDNVAKVNDSWYERFLDVENEELSR--- 1274
1115 LIATTHADIMHLD SVVRQLPLDYLDGTAL-----LVPA SVN 1155
1275 --ISEKASIEH--EALVLEADLEVQTEKLCLEKONENKQKIVCLLEBELSVVTSERN 1329
1156 CLR-----LGSWKCTLLILLRQWK--RELGS----- 1180
1330 QLRGELODMSKKTALNQLQSLSKMKKEKTOELESHEQSECLHCTQVAEAYKEXTELLQTLSS 1389
1181 -VDEILGLPELIEGV--LQADQOLMEKTKAVSFAPITVLQMKEMKYSV-----IPQ 1230
1390 DVSEILKDKTKHLOKQLQSLKDSQALSTYKCELLENQIALQNLKEREKLLVYKESSELSQARISE 1449
1231 YSOLVANCETLQOEIVATLPQTRHSLALSGATSDKOSMENDDCSRGRHQRQDGVCTLG 1290
1450 SDYERKLVNSKALE--AALVKEGFALRLSTQSEVHQLAR-GIEKLVKLEADB--KKQ 1503
1291 LHLAKELCEVDEDDSW-----LOVY--RLPIL-----PTLLTTLB--- 1325

```



```

Db 874 000EKS P V N T K S L F R E L Y S F A L H P N A F R L ----- G A S I A F N N I Y R E F ----- 916
Qy 701 S T O S G I V P C W M F V L K E M L P S Y H K R R V N S H G V R E O I G ----- C I L E I L I A I L 748
Db 917 - R E E S L V E Q P V F --- E A L V T Y M E S L A H A H D E K S L G T T O C C D A I D H L S L I E K H S L 972
Qy 749 N L C H E T D L S H S T P S L Q F I C I S L A Y T A G Q T V I N I M I G V D T I D M W A A Q P R S D G A G Q 808
Db 973 N K A K R R L P R G P P A T S - L C L ----- L D V O W L L A N G A R C P T E C R H 1012
Qy 809 G O G O L L I T V ----- K L A F S V T N N V I R L K P S N V S P L E Q A L S G H A G N N L A V A L A K - 861
Db 1013 K S I E L F Y K V T L L P G N K S P F L M K D I I K E D I S F L I N T F E ----- G G S G R P S G I L A Q P 1066
Qy 862 Y I Y H K D A L P R L A Q L K R L A T V A P M S V V A C L G D A A I A I D A F I T R Q S K - I E M R K V 920
Db 1067 T L F H L O G F S L R A A L Q M M D M L ----- L A A L E C Y N T F I E E K T L E A P K V - 1108
Qy 921 M I L E F L V A V E T Q P G L --- I E L F L - N L E V K D S D G S K E F S L G M S C L A V L E L I D S Q O D 976
Db 1109 ----- L G T E T O S L M K A V A F L E S I A M H D I M A E K I F G T G ----- A T G N R P S P O G E 1155
Qy 977 R Y --- W C P P L L R A A I A F L A L M O D R D S A M L V I R T K P F M E ----- N L T S P L F G T 1024
Db 1156 R Y N Y S K C T I V R - I M E F T T L ----- L S T S P E G W K L E K O V C N T N L M K L A V K T 1202
Qy 1025 L S P S E T S ----- E P S I L E T C A L I M K I C L E I Y Y V K S L D S L K D T L K K P S I E 1073
Db 1203 L C E P S I G F N I G D V A V M N Y L P S V - - C T N L M K A L K K S P Y - - K D I L E M L K E K I T A O S I E 1256
Qy 1074 K R P A W S G V K S L A V H A V A E T E G S C T S L E Y O M L V S A W M L - - - - - L I A T T H A D I M H 1126
Db 1257 E L C A - - - - - V D I X P D - - - - - A C V D R A R L A S V S A C K O L H R A G V L C V I I P S O S A D Q H N 1304
Qy 1127 L T D S V R R Q L P L D V L D G T K A L L L P A S V N C L R L G S M K C T L L L I L R Q W K R E I G S V D E I L G 1186
Db 1305 S I G T L L S L V Y K S I A P G D E Q C L P S L D P N C K R L A G L E L A F - - - - - A F G 1349
Qy 1187 P L T E L L E V L Q A D Q M E K T A K V S A P I T Y L Q W - - - - - K E M K V S D I P O Y S O L V I N V C E T 1241
Db 1350 G L C E H L V S L L - - - - - L D T V L S M P S R G S O K N I V S F S H G E Y F Y S L F S E T 1393
Qy 1242 L O E E V I A L F D Q T R H S I A L G S A T E D K D S M E T - - - - - D D C S R S R H D O R D G V C V L G L H A 1294
Db 1394 I N T E L L K N L D L A - - V L E L M K S V D N P K V S N V L N G M L D O S P R D T S E K H Q - - - - - G L K I A 1446
Qy 1295 K E L C E V D E D G D S W - - - - - L Q V T R R L P - - - - - I L P T L L T T L E V S L 1328
Db 1447 T I I Q N W K K C D S W M A K O S A P E S K M A V L T L A K I F O I D S V C F N T N H C M F P E V F T T Y - V S L 1505
Qy 1329 - - R M K O N H F E A T I L L T L A R T O G A T A V A G A I T O - S I C L P - - L S V I Q L S T N - - - - 1379
Db 1506 L A D S R L D L H L K G O A I T L L P F F - - - - - T S L G S L E D L K V L E N L I V S N P M K S E B P P 1558
Qy 1380 G T A O T P S - - - - - A S R K S L D A P S W G V Y R L S M S - - - - - I M E O L K T L R Y N F L P E A L D F V G H Q 1431
Db 1559 G T L O Q N N V Y D C M K K F L D A - - - - - L E L S X S P M L O L M T E I L C R E O Q H M E E L P O S T F K K I 1612
Qy 1432 E R T L O C L A V R T V O S I A C L E A D H T V G F I L O L S N F M K E - - - - - W H F H L P O L - - - 1477
Db 1613 A R K S C I T Q L G L E S V Y M F R R D D L - - - - - L S N I T R Q A F V D R S L L T L L M H C S L A N A L R E F 1666
Qy 1478 - - - - - M D I O V N L G Y L C O A C T S L H S R K M L O H Y I Q N K G D G L P S A V A Q R V O R 1524
Db 1667 F S K I V E A I N V L K S F I K L N E S A F D T O I T K M G Y Y K M L D - - - - - 1705
Qy 1525 P P S A S A P S S S K O P A D T E A S E O A L H T V O Y G L L - - - - - K I L S K T I A L R H - - F T P D V C - 1577
Db 1706 - - - - - V M S R L P K D V H S K E S K - I N Q V H G S C I T G S E L T K I L I C L Y D A F T E M A G 1756
Qy 1578 - - Q I L L D Q S L - D L A E Y N - - - - - F L P A L S F T P T P D S E V A P S F G L L A 1616
Db 1757 E N Q L L E R R L Y H C A Y N C A I S V C V F N E L K F Y O G F L F T - - - - - E K E K N I L I F E N L I D L K R 1813

```

```

Qy 1617 T V N V A L N M L G E L D K K K E P L - - - - - T O A V G L S T O A E G T R T L K S L M F T W E N C F Y L L I S Q A M 1671
Db 1814 C Y T P F I E V E W M E R K K T L E I R K E A R E A A A S G D S D G F R Y I S L - - - - - 1856
Qy 1672 R Y L R D P A V H P R D K O R M Q O E L S - E L S T L S L S R Y F R R G A P S S P A T G V L P S P O G S T S L S 1730
Db 1857 S Y L A D S - - - - - L S E M S Q P D F S T G V Q S Y S - - - - - Y S S Q D P K S T T A H 1893
Qy 1731 K A S P S Q E P L I O - - - - - L V O A F R H M O R 1753
Db 1894 F R O Q H K E S M I Q D D I L E M D E L N Q H E C M A T M T A L I K H M O R 1934

```

Search completed: December 30, 2004, 15:15:55  
 Job time : 113.809 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2004, 15:26:41 ; Search time 191 Seconds  
(without alignments)  
42.174 Million cell updates/sec

Title: US-10-719-385-20  
Perfect score: 14  
Sequence: 1 MIRKSTSVLSFC 14

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 150 summaries

Database : Uniprot 02:.\*  
1: uniprot\_sprot:.\*  
2: uniprot\_trembl:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	8	57.1	0821R8	Q821R8 chlamydom
2	7	50.0	92 1 VB41 STRMU	P59154 streptococ
3	7	50.0	145 2 Q7Y18	Q7Y18 cryptospori
4	7	50.0	347 2 Q7N4K4	Q7N4K4 photorhabd
5	7	50.0	431 2 Q831Y4	Q831Y4 enterococc
6	7	50.0	447 2 Q81G21	Q81G21 bacillus ce
7	7	50.0	696 2 Q8EN85	Q8EN85 oceanobacil
8	7	50.0	987 2 Q6BN14	Q6BN14 debaryomyce
9	7	50.0	1018 2 Q8FME0	Q8FME0 corynebacte
10	7	50.0	1019 2 Q8NM51	Q8NM51 corynebacte
11	7	50.0	1019 2 CAF20751	CAF20751 corynebac
12	7	50.0	1027 2 Q6CIE3	Q6CIE3 kluyveromyc
13	7	50.0	1601 2 Q45231	Q45231 caenorhabdi
14	6	42.9	43 2 Q48626	Q48626 leuconostoc
15	6	42.9	56 2 Q8ZIC7	Q8ZIC7 yersinia pe
16	6	42.9	56 2 AAS63080	AAS63080 yersinia
17	6	42.9	81 2 Q9P1F0	Q9P1F0 homo sapien
18	6	42.9	81 2 P70991	P70991 bacillus su
19	6	42.9	84 2 Q8KVS0	Q8KVS0 bacillari
20	6	42.9	101 2 Q8KHV2	Q8KHV2 renibacteri
21	6	42.9	105 2 Q7Q7E6	Q7Q7E6 glaciaria lam
22	6	42.9	126 2 Q918L5	Q918L5 brachydanio
23	6	42.9	131 2 Q7UDV4	Q7UDV4 rhodospirill
24	6	42.9	143 1 RL15 METVA	RL15 METVA
25	6	42.9	143 2 Q6LXDI	Q6LXDI methanococ
26	6	42.9	143 2 CAF30977	CAF30977 methanoc
27	6	42.9	153 2 Q7VC75	Q7VC75 prochloroco
28	6	42.9	157 2 Q7RKH6	Q7RKH6 plasmodium
29	6	42.9	167 2 Q737A5	Q737A5 bacillus ce
30	6	42.9	167 2 AAS41657	AAS41657 bacillus
31	6	42.9	169 2 Q70GJ2	Q70GJ2 pyrus commu

32	6	42.9	169 2 Q70GJ3	Q70GJ3 pyrus commu
33	6	42.9	169 2 CAE46475	CAE46475 pyrus com
34	6	42.9	169 2 CAE46476	CAE46476 pyrus com
35	6	42.9	170 2 Q70G16	Q70G16 pyrus commu
36	6	42.9	170 2 CAE46655	CAE46655 pyrus com
37	6	42.9	173 2 Q9L362	Q9L362 yersinia ps
38	6	42.9	173 2 Q9L364	Q9L364 yersinia pe
39	6	42.9	174 2 Q6H108	Q6H108 bacillus th
40	6	42.9	174 2 Q81PS9	Q81PS9 bacillus an
41	6	42.9	174 2 RAT31833	RAT31833 bacillus
42	6	42.9	181 2 Q7P8G7	Q7P8G7 fusobacteri
43	6	42.9	184 1 TRIS_RABIT	TRIS_RABIT
44	6	42.9	184 1 TRIS_MOUSE	TRIS_MOUSE
45	6	42.9	186 1 TRIS_RAT	TRIS_RAT
46	6	42.9	188 2 Q90350	Q90350 colutrix co
47	6	42.9	189 2 Q98DC8	Q98DC8 rhizobium 1
48	6	42.9	192 2 Q98MW2	Q98MW2 rhizobium 1
49	6	42.9	194 2 Q9CSW0	Q9CSW0 mus musculu
50	6	42.9	196 2 Y025_BUCBP	Y025_BUCBP
51	6	42.9	196 2 Q8ZVA7	Q8ZVA7 soybean dwa
52	6	42.9	200 2 Q08391	Q08391 soybean dwa
53	6	42.9	200 2 Q870J3	Q870J3 soybean dwa
54	6	42.9	209 2 Q9EMN9	Q9EMN9 ansacta moo
55	6	42.9	219 2 Q85JY9	Q85JY9 protopelago
56	6	42.9	222 2 Q8GKJ7	Q8GKJ7 arabidopsis
57	6	42.9	229 2 Q6FZK3	Q6FZK3 bartonella
58	6	42.9	229 2 Q6BMZ9	Q6BMZ9 bartonella
59	6	42.9	234 2 Q6BMZ9	Q6BMZ9 debaryomyce
60	6	42.9	241 2 Q6F117	Q6F117 mesoplasma
61	6	42.9	242 2 Q7V3B3	Q7V3B3 prochloroco
62	6	42.9	251 2 Q8AV61	Q8AV61 brachydanio
63	6	42.9	266 2 Q7MAF5	Q7MAF5 wolinnella s
64	6	42.9	276 2 Q92R22	Q92R22 rhizobium m
65	6	42.9	286 2 Q76CE1	Q76CE1 aspergillus
66	6	42.9	286 2 Q8GYJ3	Q8GYJ3 arabidopsis
67	6	42.9	286 2 Q94CM0	Q94CM0 arabidopsis
68	6	42.9	286 2 BAD01559	BAD01559 aspergill
69	6	42.9	287 2 Q6DIR9	Q6DIR9 xenopus tro
70	6	42.9	289 2 Q48846	Q48846 arabidopsis
71	6	42.9	296 2 Q9D2A4	Q9D2A4 mus musculu
72	6	42.9	298 2 Q73GB7	Q73GB7 wolbachia p
73	6	42.9	298 2 Q7NBU4	Q7NBU4 mycoplasma
74	6	42.9	298 2 AAS14699	AAS14699 wolbachia
75	6	42.9	308 2 Q8RGM6	Q8RGM6 fusobacteri
76	6	42.9	310 2 Q8A1P5	Q8A1P5 bacterioides
77	6	42.9	311 2 Q8D2_HUMAN	Q8D2_HUMAN
78	6	42.9	312 2 Q9DFC1	Q9DFC1 xenopus lae
79	6	42.9	313 2 Q7SBL1	Q7SBL1 neurospora
80	6	42.9	316 2 Q6GBL8	Q6GBL8 streptoloco
81	6	42.9	316 2 Q6GJ58	Q6GJ58 streptoloco
82	6	42.9	316 2 Q8NXT5	Q8NXT5 streptoloco
83	6	42.9	316 2 Q9W02	Q9W02 streptoloco
84	6	42.9	316 2 Q7A738	Q7A738 streptoloco
85	6	42.9	317 2 Q6IK06	Q6IK06 drosophila
86	6	42.9	323 2 Q91BR2	Q91BR2 fuqu rubrip
87	6	42.9	324 1 DLH1_CANAL	DLH1_CANAL
88	6	42.9	328 2 Q6HT8	Q6HT8 bacillus th
89	6	42.9	328 2 Q736E1	Q736E1 bacillus ce
90	6	42.9	328 2 AAS41871	AAS41871 bacillus
91	6	42.9	331 1 FH4_HUMAN	FH4_HUMAN
92	6	42.9	340 1 PPTA_ARATH	PPTA_ARATH
93	6	42.9	343 2 Q9RZB5	Q9RZB5 deinozococcus
94	6	42.9	366 2 Q98441	Q98441 eusigmamos
95	6	42.9	366 2 Q98442	Q98442 opioocytium
96	6	42.9	368 2 Q8SRX4	Q8SRX4 encaphalito
97	6	42.9	385 2 Q9XF11	Q9XF11 arabidopsis
98	6	42.9	388 2 Q8SRZ8	Q8SRZ8 encaphalito
99	6	42.9	389 2 Q7V386	Q7V386 prochloroco
100	6	42.9	396 2 Q9FR73	Q9FR73 lepidium ca
101	6	42.9	401 2 Q96AJ1	Q96AJ1 homo sapien
102	6	42.9	401 2 Q75138	Q75138 homo sapien
103	6	42.9	403 2 Q8DY18	Q8DY18 streptococ
104	6	42.9	403 2 Q8E3N7	Q8E3N7 streptococ

ID	Q821R8	PRELIMINARY	PRT	506 AA.
105	6	42.9	404	2
106	6	42.9	404	2
107	6	42.9	404	2
108	6	42.9	404	2
109	6	42.9	404	2
110	6	42.9	405	2
111	6	42.9	405	2
112	6	42.9	405	2
113	6	42.9	412	2
114	6	42.9	412	2
115	6	42.9	414	2
116	6	42.9	429	2
117	6	42.9	429	2
118	6	42.9	433	1
119	6	42.9	433	1
120	6	42.9	438	2
121	6	42.9	438	2
122	6	42.9	447	2
123	6	42.9	447	2
124	6	42.9	447	2
125	6	42.9	447	2
126	6	42.9	447	2
127	6	42.9	450	2
128	6	42.9	450	2
129	6	42.9	452	2
130	6	42.9	468	2
131	6	42.9	469	1
132	6	42.9	479	2
133	6	42.9	479	2
134	6	42.9	482	2
135	6	42.9	480	2
136	6	42.9	493	2
137	6	42.9	519	2
138	6	42.9	519	2
139	6	42.9	526	2
140	6	42.9	530	2
141	6	42.9	530	2
142	6	42.9	539	2
143	6	42.9	540	1
144	6	42.9	552	2
145	6	42.9	552	2
146	6	42.9	558	2
147	6	42.9	559	2
148	6	42.9	560	2
149	6	42.9	561	2
150	6	42.9	566	1

## ALIGNMENTS

RESULT 1

Q821R8 PRELIMINARY PRT 506 AA.

AC Q821R8: 01-JUN-2003 (TRENBLREL. 24, Created)

DT 01-JUN-2003 (TRENBLREL. 24, Last sequence update)

DT 01-MAR-2004 (TRENBLREL. 26, Last annotation update)

DE 60 kDa chaperonin, putative.

GN OrderedLocNames=CCA00870;

OS Chlamydomonas reinhardtii.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomonadales.

OX NCBI\_TaxID=83557;

RA NCBI\_TaxID=83557;

RC STRAIN=GPIC;

RX MEDLINE=22569155; PubMed=12682364;

RA Read T.D., Myers G.S.A., Brubaker R.C., Nelson W.C., Paulsen I.T.,

RA Heidelberg U.F., Holtzapfel E.K., Knouri H.M., Federova N.B.,

RA Carthy H.A., Umayam U.A., Haft D.H., Peterson J.D., Beanan M.J.,

RA White O., Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G.,

RA Bovolenta P.M., Fraser C.M.,

RT "Genome sequence of Chlamydomonas reinhardtii (Chlamydomonas reinhardtii) GPIC):

RT examining the role of niche-specific genes in the evolution of the

RT Chlamydiaceae."

RT Nucleic Acids Res. 31:2134-2147(2003).

CC -1- FUNCTION: Prevents misfolding and promotes the refolding and

CC proper assembly of unfolded polypeptides generated under stress

CC conditions (By similarity).

CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of

CC 7 subunits (By similarity).

CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.

DR EMBL: A016997; A050511.1; -.

DR HSP60; P45746; ISRV.

DR TIGR: CCA00870; -.

DR GO: GO:0005524; F:ATP binding; IEA.

DR GO: GO:0003754; F:Chaperone activity; IEA.

DR GO: GO:0006457; F:Protein folding; IEA.

DR InterPro: IPR001844; Chaperonin Cpn60.

DR InterPro: IPR002423; Cpn60/TCP-1.

DR InterPro: IPR008950; GroEL-ATPase.

DR Pfam: PF00118; Cpn60\_TCP1.1.

DR PRINTS: PR00298; CHAPERONIN60.

DR ATP-binding; Chaperone; Complete proteome.

SW SEQUENCE 506 AA; 55976 MW; A57144199770239A CRC64;

Query Match 57.1%; Score 8; DB 2; Length 506;

Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KITSVLSF 13

Db 213 KITSVLSF 220

RESULT 2

YB41 STRMU STANDARD; PRT; 92 AA.

AC P59154; 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 01-OCT-2004 (Rel. 45, Last annotation update)

DE Hypothetical UPF0223 protein SMU.1141c.

GN OrderedLocNames=SMU.1141c;

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RA NCBI\_TaxID=1309;

RC STRAIN=UAI59 / ATCC 700610 / Serotype C;

RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;

RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,

RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lan S.P.,

RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,

RA Ferretti J.J.,

RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental

RT pathogen."

RU Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

CC -1- SIMILARITY: Belongs to the UPF0223 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: A014951; AAN58833.1; -.

DR HAMAP: MF\_01041; -; 1.

DR InterPro: IPR007920; UPF0223.

DR Pfam: PF05256; UPF0223; 1.

KM Complete proteome; Hypothetical protein.

SW SEQUENCE 92 AA; 10622 MW; B8C844A96FE0BD1 CRC64;

Query Match 50.0%; Score 7; DB 1; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ITSVLSF 13  
 |||||  
 18 ITSVLSF 24

## RESULT 3

ID Q7YY18 PRELIMINARY; PRT; 145 AA.

AC Q7YY18  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DE Hypothetical protein (Fragment).  
 GN ORFNames=1MB.86;  
 OS Cryptosporidium parvum.  
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
 OC Cryptosporidiidae; Cryptosporidium.  
 NCBI\_TaxID=5807;  
 RX NCI [1]

RP SEQUENCE FROM N.A.  
 RA Bankier A.T., Spriggs H.F., Partmann B., Konfortov B.A., Madera M.,  
 RA Vogel C., Teichmann S.A., Ivens A., Dear P.H.;  
 RT "Integrated mapping, chromosomal sequencing and sequence analysis of  
 RT Cryptosporidium parvum."  
 RL Genome Res. 0:0-0(2003).  
 DR EMBL; BX538350; CAD98670.1; -.  
 KM Hypothetical protein.  
 FT NOW TRR 145 145

SO SEQUENCE 145 AA; 16639 MW; 36C9CF68CE2B299 CRC64;  
 Query Match 50.0%; Score 7; DB 2; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SKITSVL 11  
 |||||  
 27 SKITSVL 33

## RESULT 4

ID Q7NAK4 PRELIMINARY; PRT; 347 AA.

AC Q7NAK4  
 DT 01-MAR-2004 (TREMBlrel. 26, Created)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Similar to unknown protein y1ah of Escherichia coli.  
 GN OrderedLocustNames=plu2327;  
 OS Photorhabdus luminescens (subsp. laumondii).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Photorhabdus.  
 NCBI\_TaxID=141679;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=T01;  
 RX MEDLINE=22957627; PubMed=14528314;  
 RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,  
 RA Taouiri S., Bocs S., Bouraux-Bude C., Chandler M., Charles J.-F.,  
 RA Daes E., Derose R., Derzelle S., Freysinet G., Gaudault S.,  
 RA Medigue C., Lanois A., Powell K., Sigulier P., Vincent R., Wingate V.,  
 RA Zouine M., Glaser P., Boemare N., Danchin A., Kunet F.;  
 RT "The genome sequence of the entomopathogenic bacterium Photorhabdus  
 RT luminescens".  
 RL Nat. Biotechnol. 21:1307-1313(2003).

DR EMBL; BX571866; CAB14620.1; -.  
 DR Photoblast; plu2327; -.  
 DR GO; GO:0016747; F:transferase activity, transferring groups O. .; IEA.  
 DR InterPro; IPR002656; Acyl\_transf\_3.  
 DR Pfam; PF01757; Acyl\_transf\_3; 1.

KW Complete proteome.  
 SO SEQUENCE 347 AA; 40078 MW; EB4C7F7429C3C299 CRC64;  
 Query Match 50.0%; Score 7; DB 2; Length 347;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RKSKITS 9  
 |||||  
 336 RKSKITS 342

## RESULT 5

ID Q831Y4 PRELIMINARY; PRT; 431 AA.

AC Q831Y4  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Adenylsuccinate lyase.  
 GN Name=purB; OrderedLocustNames=EF2361;  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 OC Bacteria; Firmicutes; Lactobacilliales; Enterococcaceae; Enterococcus.  
 NCBI\_TaxID=1351;  
 RX NCI [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=V563 / ATCC 700802;  
 RX MEDLINE=22550857; PubMed=12663927; DOI=10.1126/science.1080613;  
 RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seasholtz R.,  
 RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,  
 RA Tettelin H., Dodson R.J., Umayam L.A., Brinkac L.M., Beanan M.J.,  
 RA Dougherty S.C., Deboy R.T., Durkin S.A., Kolonay J.F., Madupu R.,  
 RA Nelson W.C., Vamathevan J.J., Tran B., Upton J., Hansen T., Shetty J.,  
 RA Khouri H.M., Ueberback T.R., Radune D., Ketchum K.A., Dougherty B.A.,  
 RA Fraser C.M.;  
 RT "Role of mobile DNA in the evolution of vancomycin-resistant  
 RT Enterococcus faecalis".  
 RL Science 299:2071-2074(2003).  
 DR EMBL; AE016954; AA082083.1; -.  
 DR HSSP; Q9X010; IC3C.

DR TIGR; EF2361; -.  
 DR GO; GO:0004018; F:adenylsuccinate lyase activity; IEA.  
 DR GO; GO:0016829; F:lyase activity; IEA.  
 DR GO; GO:0009152; P:purine ribonucleotide biosynthesis; IEA.  
 DR InterPro; IPR003031; D\_Crystallin.  
 DR InterPro; IPR003062; Fumarate\_lyase.  
 DR InterPro; IPR008948; L-Asparagine-like.  
 DR InterPro; IPR004769; Pur\_lyase.  
 DR Pfam; PF00206; Lyase 1; 1.  
 DR PRINTS; PRO0145; DCRYSTALLIN.  
 DR PRINTS; PRO0149; FUMARATELYASE.  
 DR TIGRFAMs; TIGR00928; purB; 1.  
 DR PROSITE; PS00163; FUMARATE\_LYASES; 1.  
 KW Complete proteome; Lyase.  
 SO SEQUENCE 431 AA; 49171 MW; 5F60A97EA10F781A CRC64;

Query Match 50.0%; Score 7; DB 2; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KITSVLS 12  
 |||||  
 398 KITSVLS 404

## RESULT 6

ID Q81G21 PRELIMINARY; PRT; 447 AA.

AC Q81G21  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Potassium uptake protein KtrB.



GN ORFNames=BC1390;  
 OS *Bacillus cereus* (strain ATCC 14579 / DSM 31).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OK NCBI\_TaxID=226900;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;  
 RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,  
 RA Kapriel V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,  
 RA Chu L., Mazur M., Goldsman E., Larsen N., D'Souza M., Walunas T.,  
 RA Grechkin Y., Busch G., Haselkorn R., Fomstein M., Ehrlich S.D.,  
 RA Overbeek R., Kyriides N.C.;  
 RT "Genome sequence of *Bacillus cereus* and comparative analysis with  
 RT *Bacillus anthracis*."  
 RL Nature 423:87-91(2003).  
 DR EMBL, AF017002; AAP08372.1; -.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0008324; F:cation transporter activity; IEA.  
 DR GO; GO:0006812; P:cation transport; IEA.  
 DR InterPro; IPR003445; Cat transpt.  
 DR Pfam; PF02386; T-kt; 1.  
 DR TIGRFAMs; TIGR00933; 2a38, 1.  
 SQ SEQUENCE 447 AA; 49036 MW; AA80200421EB1199 CRC64;  
 Query Match 50.0%; Score 7; DB 2; Length 447;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 ITSVSF 13  
 DB 325 ITSVSF 331  
 RESULT 7  
 QGEN85 PRELIMINARY; PRT; 696 AA.  
 AC QGEN85;  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Transcriptional antiterminaltor (BglG family).  
 GN OrderedLocuNames=OB2602;  
 OS *Oceanobacillus theysensis*.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.  
 OK NCBI\_TaxID=182710;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=HTE831;  
 RX MEDLINE=22220767; PubMed=12235376;  
 RA Takami H., Takaki Y., Uchiyama I.;  
 RT "Genome sequence of *Oceanobacillus theysensis* isolated from the Iheya  
 RT Ridge and its unexpected adaptive capabilities to extreme  
 RT environments."  
 RL Nucleic Acids Res. 30:3927-3935(2002).  
 DR EMBL, AF004601; BAC14558.1; -.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003531; F:sugar porter activity; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. .; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR001550; BglG\_antitermin.  
 DR InterPro; IPR009059; bi\_resp\_regltr\_C.  
 DR InterPro; IPR002197; HTH\_Fis.  
 DR InterPro; IPR003501; Ptrans\_IIB.  
 DR InterPro; IPR002178; Pts\_EIIA\_2.  
 DR InterPro; IPR009058; wing\_hlx\_DNA\_bnd.  
 DR Pfam; PF00874; FRD; 2.  
 DR Pfam; PF02302; Pts\_IIB; 1.  
 DR ProDom; PD001689; Pts\_EIIA\_2; 1.  
 DR TIGRFAMs; TIGR01199; HTH\_fis; 1.  
 KM Complete proteome.

SEQ SEQUENCE 696 AA; 80117 MW; SE7DBD0075115CD CRC64;  
 Query Match 50.0%; Score 7; DB 2; Length 696;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 RKSITS 9  
 DB 500 RKSITS 506  
 RESULT 8  
 QGEN14 PRELIMINARY; PRT; 987 AA.  
 AC QGEN14;  
 DT 01-OCT-2004 (TRENBLrel. 28, Created)  
 DT 01-OCT-2004 (TRENBLrel. 28, Last sequence update)  
 DT 01-OCT-2004 (TRENBLrel. 28, Last annotation update)  
 DE Similar to ap|P38810 *Saccharomyces cerevisiae* SBD5-binding protein  
 DE 3  
 GN ORFNames=DEHA021956;  
 OS *Debaryomyces hansenii* (Yeast) (*Torulaspora hansenii*).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
 OK NCBI\_TaxID=4959;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=CS767;  
 RG GENOLEVURES;  
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neveglisse C., Talla E.,  
 RA Goffard N., Frangeul L., Aigle M., Anthonard V., Babour A., Barbe V.,  
 RA Barnay S., Blanchin S., Beckerich J.M., Bayne E., Bleykasten C.,  
 RA Boistrane A., Boyer J., Catolico L., Confiantier F., de Daruvar A.,  
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 RA Hantreaye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
 RA Kerrest A., Koszul R., Lemaire M., Leduc I., Ma L., Muller H.,  
 RA Nicoud J.M., Nikoleki M., Ozras S., Ozier-Kalogeropoulos O.,  
 RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,  
 RA Suenne D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
 RA Zenou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weisenbach J.,  
 RT "Genome evolution in Yeasts."  
 RL Nature 430:35-44(2004).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=CS767;  
 RA Genoscope;  
 RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, CR382137; CAG88479.1; -.  
 SQ SEQUENCE 987 AA; 109971 MW; BF2CCD55806DABOC CRC64;  
 Query Match 50.0%; Score 7; DB 2; Length 987;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 KITSVS 12  
 DB 461 KITSVS 467  
 RESULT 9  
 QGEN10 PRELIMINARY; PRT; 1018 AA.  
 AC QGEN10;  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Putative Na+/H+ antiporter.  
 GN OrderedLocuNames=CE3567;  
 OS *Corynebacterium efficiens*.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=152794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Y5-314;  
 RX MEDLINE=22723752; PubMed=12840036;  
 RA Nishio Y., Nakamura Y., Kawarayashi Y., Usuda Y., Kimura E.,  
 RA Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,  
 RA Gotohori T.;  
 RT "Comparative complete genome analysis of the amino acid  
 RT replacements responsible for the thermostability of Corynebacterium  
 RT efficiens.";  
 RL Genome Res. 13:1572-1579(2003).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 DR EMBL: AP005222; BAC1937.1; -  
 DR GO: GO:0016021; C:Integral to membrane; IEA.  
 DR GO: GO:0008137; P:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO: GO:0016491; P:oxidoreductase activity; IEA.  
 DR GO: GO:0042773; P:ATP synthesis coupled electron transport; IEA.  
 DR InterPro: IPR007182; MnhB.  
 DR InterPro: IPR003916; NADH\_oxrds.  
 DR InterPro: IPR001750; Oxidored\_q1.  
 DR InterPro: IPR001516; Oxidored\_q1\_N.  
 DR Pfam: PF04039; MnhB, 1.  
 DR Pfam: PF00361; Oxidored\_q1; 1.  
 DR Pfam: PF00662; Oxidored\_q1\_N; 1.  
 DR PRINTS: PR01434; NADHGNASE5.  
 DR Complete proteome; Transmembrane.  
 KW SEQUENCE 1018 AA; 110100 MW; A1SEFDAB9F36015 CRC64;  
 SQ  
 Query March 50.0%; Score 7; DB 2; Length 1018;  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 7 ITSVLSF 13  
 Db 139 ITSVLSF 145  
 RESULT 10  
 Q8NM51 PRELIMINARY; PRT; 1019 AA.  
 ID Q8NM51; Q8NM51; Q8NM51; Q8NM51;  
 AC Q8NM51; Q8NM51; Q8NM51; Q8NM51;  
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 28, Last annotation update)  
 DT 01-OCT-2004 (TREMblrel. 28, Last annotation update)  
 DE NADH:ubiquinone oxidoreductase subunit 5 (Chain L)/Multisubunit Na+/H+  
 DE antiporter, MnhA subunit (NADH ubiquinone oxidoreductase subunit 5  
 DE (Chain L)/Multisubunit Na+/H+ antiporter, A subunit).  
 GN Name:mpa; Order:edLocNames=Cg12729, Cg3024;  
 OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025 Nakagawa;  
 RA Nakagawa S.;  
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";  
 RL submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
 RX PubMed=12948626;  
 RA Kalinowski J., Bathe B., Bartels D., Bischoff N., Bott M.,  
 RA Burkovski A., Dusch N., Eggeling L., Eikmanns B.J., Gaigalat L.,  
 RA Goessmann A., Hartmann M., Huthmacher K., Kraemer R., Linke B.,  
 RA McHardy A.C., Meyer F., Moeckel B., Pfeifferle W., Puhler A.,  
 RA Rey D.A., Rueckert C., Rupp O., Sahn H., Wendisch V.F., Wiegand I.,  
 RA Tauch A.;  
 RT "The complete Corynebacterium glutamicum ATCC 13032 genome sequence  
 RT and its impact on the production of L-aspartate-derived amino acids  
 RT and vitamins.";

RL J. Biotechnol. 104:5-25(2003).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 DR EMBL: AP005282; BAC00123.1; -  
 DR EMBL: BX927156; CAF20751.1; -  
 DR GO: GO:0016021; C:Integral to membrane; IEA.  
 DR GO: GO:0008137; P:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO: GO:0016491; P:oxidoreductase activity; IEA.  
 DR GO: GO:0042773; P:ATP synthesis coupled electron transport; IEA.  
 DR InterPro: IPR007182; MnhB.  
 DR InterPro: IPR003916; NADH\_oxrds.  
 DR InterPro: IPR001750; Oxidored\_q1.  
 DR InterPro: IPR001516; Oxidored\_q1\_N.  
 DR Pfam: PF04039; MnhB, 1.  
 DR Pfam: PF00361; Oxidored\_q1; 1.  
 DR Pfam: PF00662; Oxidored\_q1\_N; 1.  
 DR PRINTS: PR01434; NADHGNASE5.  
 DR Complete proteome; Transmembrane; Ubiquinone.  
 KW SEQUENCE 1019 AA; 110212 MW; A7A7372E155E0B57 CRC64;  
 SQ  
 Query March 50.0%; Score 7; DB 2; Length 1019;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 7 ITSVLSF 13  
 Db 139 ITSVLSF 145  
 RESULT 11  
 CAF20751 PRELIMINARY; PRT; 1019 AA.  
 ID CAF20751; CAF20751; CAF20751; CAF20751;  
 AC CAF20751; CAF20751; CAF20751; CAF20751;  
 DT 02-MAR-2004 (TREMblrel. 27, Last sequence update)  
 DT 02-MAR-2004 (TREMblrel. 27, Last annotation update)  
 DT 02-MAR-2004 (TREMblrel. 27, Last annotation update)  
 DE NADH ubiquinone oxidoreductase subunit 5 (Chain L)/Multisubunit Na+/H+  
 DE antiporter, A subunit.  
 GN MRPA OR CG3024.  
 OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
 RX PubMed=12948626;  
 RA Kalinowski J., Bathe B., Bartels D., Bischoff N., Bott M.,  
 RA Burkovski A., Dusch N., Eggeling L., Eikmanns B.J., Gaigalat L.,  
 RA Goessmann A., Hartmann M., Huthmacher K., Kraemer R., Linke B.,  
 RA McHardy A.C., Meyer F., Moeckel B., Pfeifferle W., Puhler A.,  
 RA Rey D.A., Rueckert C., Rupp O., Sahn H., Wendisch V.F., Wiegand I.,  
 RA Tauch A.;  
 RT "The complete Corynebacterium glutamicum ATCC 13032 genome sequence  
 RT and its impact on the production of L-aspartate-derived amino acids  
 RT and vitamins.";  
 RL J. Biotechnol. 104:5-25(2003).  
 DR EMBL: BX927156; CAF20751.1; -  
 KW Ubiquinone.  
 SQ SEQUENCE 1019 AA; 110212 MW; A7A7372E155E0B57 CRC64;  
 Query March 50.0%; Score 7; DB 2; Length 1019;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 7 ITSVLSF 13  
 Db 139 ITSVLSF 145  
 RESULT 12  
 Q6CIE3 PRELIMINARY; PRT; 1027 AA.  
 ID Q6CIE3; Q6CIE3; Q6CIE3; Q6CIE3;  
 AC Q6CIE3; Q6CIE3; Q6CIE3; Q6CIE3;

Query Match	50.0%	Score 71	DB 2	Length 1027
Best Local Similarity	100.0%	Pred. No. 96		
Matches	71	Conservative	0	Mismatches 0
				Indels 0
				Gaps 0
QY	3	RKSITS	9	
Db	497	RKSITS	503	

RESULT 13			
045231	ID	045231	PRELIMINARY; PRT; 1601 AA.
AC	045231;		
DT	01-JUN-1998	(TEMBLrel. 06, Created)	
DT	01-AUG-1998	(TEMBLrel. 07, Last sequence update)	
DT	01-MAR-2004	(TEMBLrel. 26, Last annotation update)	
DE	Hypothetical protein C01B9.1.		
GN	Name=C01B9.1;		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
UN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Bristol N2;		
RC	MEDLINE=99069613;	PubMed=9851916;	
RA	none;		
RT	"genome sequence of the nematode C.elegans: A platform for investigating biology.";		
RT	Science 282:2012-2018(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Bristol N2;		
RA	Percy C.M.;		
RL	Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.		
RR	EMBL; Z93375; CAB07551.1; -.		

Query Match	50.0%	Score 7	DB 2	Length 1601
Best Local Similarity	100.0%	Pred. No.	1.4e+02	
Matches	7	Conservative	0	Mismatches 0; Indels 0; Gaps 0

RESULT 14	
Q48626	
ID Q48626	PRELIMINARY; PRT; 43 AA

RESULT 14			
Q48626			
ID	Q48626	PRELIMINARY;	PRT; 43 AA.
AC	Q48626;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, last sequence update)		
DT	01-JUN-2003 (TrEMBLrel. 24, last annotation update)		
DE	ORF6 of IS3-like IS.		
OS	Leuconostoc lactis.		
OG	Plasmid pNZ63.		
OC	Bacteria; Firmicutes; Lactobacillales; Leuconostoc.		
OX	NCBI_TaxID=1246;		
RN			
RP	[1]		
RF	SEQUENCE FROM N.A.		
RX	STRAIN=6009;		
RX	MEDLINE=96209221; PubMed=863385;		
RA	Vaughan E.E., David S., de Vos W.M.;		
RT	"The lactose transporter in leuconostoc lactis is a new member of the		
RT	lactac family of galactoside-pentose-hexuronide translocators.";		
RL	Appl. Environ. Microbiol. 62:1574-1582(1996).		
DR	EMBL; U47655; AAC4415.1; --		
Q4	Plasmid.		
SQ	SEQUENCE 43 AA; 5279 MW; 440E9C6E7D461689 CRC64;		

Query Match	42.9%;	Score 6;	DB 2;	length 43;
Best Local Similarity	100.0%;	Pred. No. 72;		
Matches	6;	Conservative	0;	Indels 0;
		Mismatches	0;	Gaps 0

RESULT 15	
Q8ZIC7	
ID Q8ZIC7	PRELIMINARY;
	PRT;
	56 AA

	RESULT 15
08ZIC7	PRELIMINARY; PRT; 56 AA.
ID	08ZIC7
AC	08ZIC7; Q7ARW6;
DT	01-MAR-2002 (TrEMBLrel. 20, Created)
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT	01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE	Hypothetical protein YP00578.
GN	OrthoducsinNames=YP2898, YP00578;
OS	Yersinia pestis.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC	Enterobacteriaceae; Yersinia.
OX	NCBI_TaxID=632;
ON	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CO-92 / Biovar Orientalis;
RC	MEDLINE=21470413; PubMed=11586360; DOI=10.1038/35097083;
RA	Parhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
RA	Prentice M.B., Sebahia M., James K.D., Churcher C.M., Mungall K.L.,
RA	Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.-M.,
RA	Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA	Feltwell T., Hamlin N., Holroyd S., Jagers K., Kariyeh A.V.,
RA	Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.M.,
RA	Simmonds M., Skelton S., Stevens K., Whitehead S., Barrall B.G.;

RT "Genome sequence of Yersinia pestis, the causative agent of plague."  
RL Nature 413:523-527(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=91001 / Biovar Mediaevalis;  
RA Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D.,  
Han Y., Pang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,  
Yang C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang P.,  
Yang R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ414143; CAC89435.1; -  
DR EMBL; AE017138; AAS63080.1; -  
DR PIR; AH0071; AH0071.  
KM Complete proteome; Hypothetical protein.  
SQ SEQUENCE 56 AA; 6444 MW; C75BFB5B1F77AC90 CRC64;  
  
Query Match 42.9%; Score 6; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 7 ITSVLS 12  
Db 48 ITSVLS 53  
  
RESULT 16  
AAS63080 PRELIMINARY; PRT; 56 AA.  
AC AAS63080;  
DT 24-MAR-2004 (TREMBLrel. 27, Created)  
DT 24-MAR-2004 (TREMBLrel. 27, Last sequence update)  
DT 04-MAY-2004 (TREMBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
GN YP2898.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxId=632;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=91001 / Biovar Mediaevalis;  
RA Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D.,  
Han Y., Pang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,  
Yang C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang P.,  
Yang R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE017138; AAS63080.1; -  
KM Hypothetical protein.  
SQ SEQUENCE 56 AA; 6444 MW; C75BFB5B1F77AC90 CRC64;  
  
Query Match 42.9%; Score 6; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 7 ITSVLS 12  
Db 48 ITSVLS 53  
  
RESULT 17  
09P1F0 PRELIMINARY; PRT; 81 AA.  
AC 09P1F0;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE PRO2049.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE=Liver;  
RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,  
Li M., He F.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF116685; AAF71105.1; -  
SQ SEQUENCE 81 AA; 9410 MW; 862BF81C575F6DED CRC64;  
  
Query Match 42.9%; Score 6; DB 2; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 IRKSKI 7  
Db 42 IRKSKI 47  
  
RESULT 18  
P70991 PRELIMINARY; PRT; 81 AA.  
ID P70991;  
AC P70991;  
DT 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Hypothetical protein (Fragment).  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxId=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA MEDLINE=87080286; PubMed=3098560;  
RX Parsot C.;  
RT "Evolution of biosynthetic pathways: a common ancestor for threonine  
synthase, threonine dehydratase and D-serine dehydratase."  
RL EMO J. 5:3013-3019(1986).  
DR EMBL; X04603; CAA28269.1; -  
DR GO; GO:0016597; F:amino acid binding; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR002912; ACT.  
DR Pfam; PF01842; ACT; 1.  
KM Hypothetical protein.  
FT NON TER  
SQ SEQUENCE 81 AA; 9111 MW; FA60AB27A08FB3F0 CRC64;  
  
Query Match 42.9%; Score 6; DB 2; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 5 SKITSV 10  
Db 11 SKITSV 16  
  
RESULT 19  
O8KVS0 PRELIMINARY; PRT; 84 AA.  
AC O8KVS0;  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Putative early growth response protein.  
GN Name=grp;  
OS Renibacterium salmoninarum.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Micrococcales; Micrococcaceae; Renibacterium.  
OX NCBI\_TaxId=1646;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIMB 1114;  
RX MEDLINE=22043279; PubMed=12047757;  
RA Grayson T.H., Cooper L.F., Wratheill A.B., Roper J., Evenden A.J.,  
Gilpin M.L.;  
RT "Host responses to Renibacterium salmoninarum and specific components

RT of the pathogen reveal the mechanisms of immune suppression and  
 RT activation.";  
 RL Immunology 106:273-283(2002).  
 DR EMBL: AF428072; AAM47191.1;  
 SO SEQUENCE 84 AA; 8506 MW; 55D41CBABCE311F1 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSFC 14  
 Db 63 SVLSFC 68

## RESULT 20

08KHV2 PRELIMINARY; PRT; 101 AA.

AC 08KHV2; 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DE 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Putative early growth response protein.

GN Name=grp;  
 OS Renibacterium salmoninarum.  
 OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;  
 OC Micrococcineae; Micrococcaceae; Renibacterium.  
 CX NCBI\_TaxID=1646;

RT by random amplified polymorphic DNA analysis."

RT Appl. Environ. Microbiol. 66:435-438(2000).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=ATCC33209;  
 RC MEDLINE=20087576; PubMed=10618262;  
 RX Grayson T.H., Atlenzar F.A., Alexander S.M., Cooper L.F., Gilpin M.L.;  
 RA "Molecular diversity of Renibacterium salmoninarum isolates determined  
 RT by random amplified polymorphic DNA analysis."

RA Appl. Environ. Microbiol. 66:435-438(2000).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=ATCC33209;  
 RC MEDLINE=20043279; PubMed=12047757;  
 RX Grayson T.H., Cooper L.F., Wrathmell A.B., Roper J., Evenden A.J.,  
 RA Gilpin M.L.;  
 RA "Host responses to Renibacterium salmoninarum and specific components  
 RT of the pathogen reveal the mechanisms of immune suppression and  
 RT activation."

RL Immunology 106:273-283(2002).  
 DR EMBL: AF178994; AAM45390.1;  
 DR EMBL: AF428071; AAM47188.1;  
 SO SEQUENCE 101 AA; 10020 MW; 4D92137221251FCT CRC64;

Query Match 42.9%; Score 6; DB 2; Length 101;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSFC 14  
 Db 80 SVLSFC 85

## RESULT 21

07OT66 PRELIMINARY; PRT; 105 AA.

AC 07OT66; 01-MAR-2004 (TREMBlrel. 26, Created)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)  
 DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE GEP 13 5047 4730.

OS Giardia lamblia ATCC 50803.  
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.  
 CX NCBI\_TaxID=184922;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MB C6;

RA Morrison H.G., McArthur A.G., Adam R.D., Aley S.B., Gillin F.D.,  
 RA Olsen G.U., Sogin M.L.;  
 RT "Draft sequence of the Giardia lamblia genome."  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.

DR EMBL: AACB01000110; EAA38197.1;  
 SO SEQUENCE 105 AA; 11562 MW; F2D35D3E9564AAB8 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKITS 9  
 Db 41 KSKITS 46

## RESULT 22

0918L5 PRELIMINARY; PRT; 126 AA.

AC 0918L5; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DE 01-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 DE Liver-basis fatty acid binding protein (Hypothetical protein).

GN Name=fabp10;  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 CX Cyprinidae; Danio.  
 CX NCBI\_TaxID=7955;

RT by random amplified polymorphic DNA analysis."  
 RL [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=20461782; PubMed=11004494;  
 RX Denovan-Wright E.M., Pierce M., Sharma M.K., Wright J.M.;  
 RA "cDNA sequence and tissue-specific expression of a basic liver-type  
 RT fatty acid binding protein in adult zebrafish (Danio rerio)."  
 RL Biochim. Biophys. Acta 1492:227-232 (2000).

RL [2]  
 RN SEQUENCE FROM N.A.  
 RP Denovan-Wright E.M., Pierce M., Sharma M.K., Wright J.M.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=22386257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.U., Uedin L.B., Toshlyuk S., Carninci P., Frange C.J.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson P., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guarnatone P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Strausberg R.;  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.

```

CC -1- SIMILARITY: Belongs to the fatty-acid binding protein (FABP)
CC family.
DR EMBL: AF546442; AAF67743.1; -.
DR EMBL: BC076219; AAH76219.1; -.
DR HSSP: P83409; 1P8P.
DR ZFIN: ZDB-GENE-020318-1; fabp10.
DR GO: GO:0008289; F:lipid binding; IEA.
DR GO: GO:0005215; F:transporter activity; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR011038; Calycin.
DR InterPro: IPR000463; Fatty_acid_BP.
DR InterPro: IPR000566; Lipocin_cytrFABP.
DR InterPro: IPR010916; TONB_Box_N.
DR Pfam: PF00061; Lipocalin; 1.
DR PRINTS: PR00178; FATTYACIDBP.
DR PROSITE: PS00214; FABP; 1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
DR Hypothetical protein; Transport.
SQ SEQUENCE 126 AA; 14004 MW; F77419F1F2489814 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 MIRSKX 6
Db 119 MIRSKX 124

RESULT 23
ID 07UDV4 PRELIMINARY; PRT; 131 AA.
AC 07UDV4;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DE Hypothetical protein.
GN OrderedlocusNames=RB11760;
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heilmann K., Rabus R.,
RA Schleuter H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1."
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
KM EMBL: BX294153; CAD79303.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 131 AA; 14022 MW; 1A3D469B981B1B10 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 131;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 7 ITSVLS 12
Db 38 ITSVLS 43

RESULT 24
ID RL15 META STRAND; PRT; 143 AA.
AC RL15 META;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
KW 50S ribosomal protein L15P.

```

```

GN Name=rpl15p;
OS Methanococcus vannielii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=2187;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90040717; PubMed=2530355;
RA Auer J., Spicker G., Beck A.;
RT "Organization and structure of the Methanococcus transcriptional unit
RT homologous to the Escherichia coli 'spectinomycin operon'.
RT Implications for the evolutionary relationship of 70 S and 80 S
RT ribosomes."
RL J. Mol. Biol. 209:21-36(1989).
CC -1- SIMILARITY: Belongs to the L15P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16720; CAA34702.1; -.
DR PIR: S05626; R6MX15.
DR InterPro: IPR001196; Ribosomal_L15.
DR Pfam: PF00256; L15; 1.
DR PROSITE: PS00475; RIBOSOMAL_L15; 1.
KW Ribosomal protein.
SQ SEQUENCE 143 AA; 15950 MW; 694468C6B3217996 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 143;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 MIRSKX 6
Db 1 MIRSKX 6

RESULT 25
ID 06LXD1 PRELIMINARY; PRT; 143 AA.
AC 06LXD1;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE L15 ribosomal protein L15.
GN Name=rpl0; OrderedlocusNames=MP1421;
OS Methanococcus maripaludis.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=39152;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S2 / L1;
RA Hendrickson E.L., Kaul R., Zhou Y., Bovee D., Chapman P., Chung J.,
RA Conway de Macario E., Dodsworth J., Gillett W., Graham D.E.,
RA Haydock A.K., Kang A., Land M.L., Levy R., Lie T.J., Major T.,
RA Moore B., Porat I., Overbeek R., Palmeri A., Rouse G.,
RA Saenphimachak C., Soll D., Whitman W.B., Larimer F.W., Olson M.V.,
RA Leigh J.A.;
RT "Complete genome sequence of the mesophilic hydrogenotrophic
RT methanogen Methanococcus maripaludis."
RT Submitted (FEB-2004) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: Belongs to the L15P family of ribosomal proteins.
DR EMBL: BX57222; CAF30977.1; -.
DR GO: GO:0003735; F:structural constituent of ribosome; IEA.
DR InterPro: IPR001196; Ribosomal_L15.
DR Pfam: PF00256; L15; 1.
DR PROSITE: PS00475; RIBOSOMAL_L15; 1.
KW Complete proteome; Ribonucleoprotein; Ribosomal protein.

```

SQ SEQUENCE 143 AA; 15951 MW; AACFF8148932FB90 CRC64;  
 Query Match 42.9%; Score 6; DB 2; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MIRKSK 6  
 Db 1 MIRKSK 6  
 RESULT 26  
 CAF30977 PRELIMINARY; PRT; 143 AA.  
 ID CAF30977;  
 AC CAF30977;  
 DT 03-MAR-2004 (TREMBlrel. 27, Last sequence update)  
 DT 03-MAR-2004 (TREMBlrel. 27, Last sequence update)  
 DT 04-MAY-2004 (TREMBlrel. 27, Last annotation update)  
 DE L5U ribosomal protein L15.  
 GN RPL0 OR MPM1421.  
 OS Methanococcus maripaludis.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanococcaceae; Methanococcus.  
 CX NCBI\_TaxID=39152;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S2 / IL;  
 RA Hendrickson E.L., Kaul R., Zhou Y., Boyse D., Chapman P., Chung J.,  
 RA Conway de Macario E., Dodsworth J., Gillett W., Graham D.E.,  
 RA Haydock A.K., Kang A., Land M.L., Levy R., Lie T.J., Major T.,  
 RA Moore B., Porat I., Overbeek R., Palmeltil A., Rouse G.,  
 RA Saenphimachak C., Soell D., Whitman W.B., Larimer F.W., Olson M.V.,  
 RA Leigh J.A.;  
 RT "Complete genome sequence of the mesophilic hydrogentrophic  
 RT methanogen Methanococcus maripaludis";  
 RL Submitted (FEB-2004) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; BX957222; CAF30977.1; -.  
 KM Ribosomal protein.  
 SQ SEQUENCE 143 AA; 15951 MW; AACFF8148932FB90 CRC64;  
 Query Match 42.9%; Score 6; DB 2; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MIRKSK 6  
 Db 1 MIRKSK 6  
 RESULT 27  
 Q7VC75 PRELIMINARY; PRT; 153 AA.  
 ID Q7VC75;  
 AC Q7VC75;  
 DT 01-OCT-2003 (TREMBlrel. 25, Created)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Predicted endonuclease.  
 GN OrderedLocustNames=Pro0867;  
 OS Prochlorococcus marinus.  
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcales;  
 OC Prochlorococcus.  
 CX NCBI\_TaxID=1219;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SARG / CCMP 1375 / SS120;  
 RX MEDLINE=22810154; PubMed=12917486;  
 RA Diferene A., Salanoubat M., Partensky F., Artiguenave F., Axmann I.M.,  
 RA Barbe V., Duprat S., Galperin M.Y., Koonin E.V., Le Gall F.,  
 RA Makarova K.S., Ostrowski M., Ozlas S., Robert C., Rogozin I.B.,  
 RA Scanlan D.J., Tandeau de Marsac N., Weissenbach J., Winkler P.,  
 RA Wolf Y.I., Hess W.R.;  
 RT "genome sequence of the cyanobacterium Prochlorococcus marinus SS120,  
 a nearly minimal oxynotrophic genome."

RL Proc. Natl. Acad. Sci. U.S.A. 100:10020-10025(2003).  
 DR EMBL; AE017163; AAP99911.1; -.  
 DR GO: GO:0004519; F:endonuclease activity; IEA.  
 DR GO: GO:0016788; F:hydrolase activity, acting on ester bonds; IEA.  
 DR GO: GO:0006139; P:nucleobase, nucleoside, nucleotide and nucl. . .; IEA.  
 DR InterPro; IPR005227; Cons\_hypoth250.  
 DR InterPro; IPR006641; YGFC.  
 DR Pfam; PF03652; UPF0081.1.  
 DR SMART; SM00732; YGFC; 1.  
 DR TIGRPFAM; TIGR00250; cons\_hypoth250; 1.  
 KW Complete proteome; Endonuclease.  
 SQ SEQUENCE 153 AA; 17000 MW; 19D3E9F78FA3CB69 CRC64;  
 Query Match 42.9%; Score 6; DB 2; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 8 TSVLSF 13  
 Db 7 TSVLSF 12  
 RESULT 28  
 Q7RKH6 PRELIMINARY; PRT; 157 AA.  
 ID Q7RKH6;  
 AC Q7RKH6;  
 DT 01-MAR-2004 (TREMBlrel. 26, Created)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Hypothetical protein.  
 GN Name=PY02925;  
 OS Plasmodium yoelii yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 CX NCBI\_TaxID=73239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17XNL;  
 RX PubMed=12368865;  
 RA Carlton J.M., Anguino S.V., Suh B.B., Kooij T.W., Perce M.,  
 RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,  
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 RA Shallow S.J., van Aken S.E., Riedmiller S.B., Feldblum T.V.,  
 RA Cho J.K., Quackenbush J., Sedegah M., Shoib A., Cummings L.M.,  
 RA Florens L., Yates F.R. III, Raine J.D., Sinden J.E., Harris M.A.,  
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
 RA van Lin L.H., Jance C.J., Waters A.P., Smith H.O., White O.R.,  
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
 RA Carucci D.J.;  
 RT "genome sequence and comparative analysis of the model rodent malaria  
 RT parasite Plasmodium yoelii yoelii";  
 RL Nature 419:512-519(2002).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AABL0100823; EAA22443.1; -.  
 DR GO: GO:0009331; C:glycerol-3-phosphate dehydrogenase complex; IEA.  
 DR GO: GO:0004367; F:glycerol-3-phosphate dehydrogenase (NAD+) a. . .; IEA.  
 DR GO: GO:0016614; F:oxidoreductase activity, acting on CH-OH gr. . .; IEA.  
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO: GO:0006072; P:glycerol-3-phosphate metabolism; IEA.  
 DR InterPro; IPR006109; NAD\_Gly3P\_C.  
 DR InterPro; IPR006168; NAD\_Gly3P\_dh.  
 DR Pfam; PF07479; NAD\_Gly3P\_dh\_C; 1.  
 DR PRINTS; PR00077; GPDHGRNASE.  
 DR Prodom; PD001278; NAD\_Gly3P\_C; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 157 AA; 17683 MW; 14E354CF507B4DD CRC64;  
 Query Match 42.9%; Score 6; DB 2; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 TSVLS 12



Db 141 ITSv1.5 146

## RESULT 29

Q737A5 PRELIMINARY; PRT; 167 AA.  
AC Q737A5;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocustNames=BCE2745;  
OS Bacillus cereus (strain ATCC 10987).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=222523;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14960714;  
RA Raeko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,  
RA Shores K.A., Fouts D.E., Tourasse N.J., Angluoni S.V., Kolonay J.F.,  
RA Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;  
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic  
adaptations and a large plasmid related to Bacillus anthracis pXOI.";  
RL Nucleic Acids Res. 32:977-988(2004).  
DR EMBL; AE017273; AAS41657.1; -.  
DR TIGR; BCE2745; -.  
DR InterPro; IPR011235; DUP UCP032285.  
DR PIRSF; PIRSF032285; UCP032285; 1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 167 AA; 19611 MW; 73539A73CFDD18B7 CRC64;

Query March 42.9%; Score 6; DB 2; Length 167;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSKIT 8  
Db 50 RSKIT 55

RESULT 30  
AAS41657 PRELIMINARY; PRT; 167 AA.  
AC AAS41657;  
DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
DT 04-MAY-2004 (TrEMBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
GN BCE2745.  
OS Bacillus cereus (strain ATCC 10987).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=222523;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14960714;  
RA Raeko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,  
RA Shores K.A., Fouts D.E., Tourasse N.J., Angluoni S.V., Kolonay J.F.,  
RA Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;  
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic  
adaptations and a large plasmid related to Bacillus anthracis pXOI.";  
RL Nucleic Acids Res. 32:977-988(2004).  
DR EMBL; AE017273; AAS41657.1; -.  
DR TIGR; BCE2745; -.  
KW Hypothetical protein.  
SQ SEQUENCE 167 AA; 19611 MW; 73539A73CFDD18B7 CRC64;

Query March 42.9%; Score 6; DB 2; Length 167;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSKIT 8  
Db 50 RSKIT 55

Db 50 RSKIT 55

## RESULT 31

Q70GJ2 PRELIMINARY; PRT; 169 AA.  
AC Q70GJ2;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Nucleotide binding site leucine-rich repeat disease resistance protein  
DE (Fragment).  
OS Pyrus communis (Pear).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eucosids I; Rosales; Rosaceae; Maloideae; Pyrus.  
OX NCBI\_TaxID=23211;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Afunian M.R.;  
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ561777; CAB46476.1; -.  
FT NON TER 1 1  
FT NON TER 1 1  
SQ SEQUENCE 169 AA; 19132 MW; 38342B7F867B351B CRC64;

Query March 42.9%; Score 6; DB 2; Length 169;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
Db 51 SKITSV 56

## RESULT 32

Q70GJ3 PRELIMINARY; PRT; 169 AA.  
AC Q70GJ3;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Nucleotide binding site leucine-rich repeat disease resistance protein  
DE (Fragment).  
OS Pyrus communis (Pear).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eucosids I; Rosales; Rosaceae; Maloideae; Pyrus.  
OX NCBI\_TaxID=23211;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Afunian M.R.;  
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ561776; CAB46475.1; -.  
FT NON TER 1 1  
FT NON TER 1 1  
SQ SEQUENCE 169 AA; 19141 MW; 33971B1C03BF9620 CRC64;

Query March 42.9%; Score 6; DB 2; Length 169;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
Db 51 SKITSV 56

## RESULT 33

CAB46475 PRELIMINARY; PRT; 169 AA.  
AC CAB46475;  
DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)

DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)  
DE Nucleotide binding site leucine-rich repeat disease resistance protein  
(Fragment).  
OS Pyrus communis (Pear).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosid I; Rosales; Rosaceae; Maloideae; Pyrus.  
OX NCBI\_TaxID=23211;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Seckel;  
RA Atunian M.R.;  
RT "Molecular approaches to improving disease resistance in apple and  
pear";  
RL Submitted (JNG-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AJ581776; CAB6475.1; -.  
FT NON\_TER 1 169  
SQ SEQUENCE 169 AA; 19141 MW; 33971B1C03BF620 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 169;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SKITSV 10  
ID 11111111  
DB 51 SKITSV 56

RESULT 34  
CAB6476 PRELIMINARY; PRT; 169 AA.  
ID CAB6476;  
AC CAB6476;  
DT 02-MAR-2004 (TREMBlrel. 27, Created)  
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)  
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)  
DE Nucleotide binding site leucine-rich repeat disease resistance protein  
(Fragment).  
OS Pyrus communis (Pear).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosid I; Rosales; Rosaceae; Maloideae; Pyrus.  
OX NCBI\_TaxID=23211;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Seckel;  
RA Atunian M.R.;  
RT "Molecular approaches to improving disease resistance in apple and  
pear";  
RL Submitted (JNG-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AJ581777; CAB6476.1; -.  
FT NON\_TER 1 169  
SQ SEQUENCE 169 AA; 19132 MW; 38342B7F867B351B CRC64;

Query Match 42.9%; Score 6; DB 2; Length 169;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SKITSV 10  
ID 11111111  
DB 51 SKITSV 56

RESULT 35  
Q70G16 PRELIMINARY; PRT; 170 AA.  
ID Q70G16;  
AC Q70G16;  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
DE Nucleotide binding site leucine-rich repeat disease resistance protein  
(Fragment).

OS Pyrus communis (Pear).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosid I; Rosales; Rosaceae; Maloideae; Pyrus.  
OX NCBI\_TaxID=23211;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Seckel;  
RA Atunian M.R.;  
RT "Molecular approaches to improving disease resistance in apple and  
pear";  
RL Submitted (JNG-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AJ581783; CAB6655.1; -.  
FT NON\_TER 1 170  
SQ SEQUENCE 170 AA; 19334 MW; 73547B4FE1B89C2F CRC64;

Query Match 42.9%; Score 6; DB 2; Length 170;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SKITSV 10  
ID 11111111  
DB 52 SKITSV 57

RESULT 36  
CAB6655 PRELIMINARY; PRT; 170 AA.  
ID CAB6655;  
AC CAB6655;  
DT 02-MAR-2004 (TREMBlrel. 27, Created)  
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)  
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)  
DE Nucleotide binding site leucine-rich repeat disease resistance protein  
(Fragment).  
OS Pyrus communis (Pear).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosid I; Rosales; Rosaceae; Maloideae; Pyrus.  
OX NCBI\_TaxID=23211;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Seckel;  
RA Atunian M.R.;  
RT "Molecular approaches to improving disease resistance in apple and  
pear";  
RL Submitted (JNG-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AJ581783; CAB6655.1; -.  
FT NON\_TER 1 170  
SQ SEQUENCE 170 AA; 19334 MW; 73547B4FE1B89C2F CRC64;

Query Match 42.9%; Score 6; DB 2; Length 170;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SKITSV 10  
ID 11111111  
DB 52 SKITSV 57

RESULT 37  
Q9L362 PRELIMINARY; PRT; 173 AA.  
ID Q9L362;  
AC Q9L362;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Aerobic ribonucleotide reductase class Ib (EC 1.17.4.1)  
(Fragment).  
GN Name=nrde;  
OS Yersinia pseudotuberculosis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=633;  
RN [1]

RP SEQUENCE FROM N.A.  
 RA Torrens E., Jordan A., Gilbert I., Karlsson M.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Margaretta K.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ286852; CAB87141.1; -  
 DR GO; GO:0005971; C:ribonucleoside-diphosphate reductase complex; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0004748; F:ribonucleoside-diphosphate reductase activity; IEA.  
 DR GO; GO:0006260; P:DNA replication; IEA.  
 DR InterPro; IPR00788; Ribonucleo\_red.  
 DR Pfam; PF02867; Ribonuc\_red\_1c; 1.  
 DR PRINTS; PR01183; RIBORDTASEM1.  
 KW Oxidoreductase.  
 FT NON\_TER 1 173  
 FT MON\_TER 173  
 SQ SEQUENCE 173 AA; 19697 MW; 33C10E3B458311A CRC64;

Query Match 42.9%; Score 6; DB 2; Length 173;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 IRKSKI 7  
 Db 43 IRKSKI 48

## RESULT 38

Q9L364 PRELIMINARY; PRT; 173 AA.

AC Q9L364; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Aerobic ribonucleotide reductase class Ib (EC 1.17.4.1) (Fragment).  
 GN Name:rdc.  
 OS Versinia pectis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Torrens E., Jordan A., Gilbert I., Karlsson M.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Margaretta K.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ286850; CAB87139.1; -  
 DR GO; GO:0005971; C:ribonucleoside-diphosphate reductase complex; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0004748; F:ribonucleoside-diphosphate reductase activity; IEA.  
 DR GO; GO:0006260; P:DNA replication; IEA.  
 DR InterPro; IPR00788; Ribonucleo\_red.  
 DR Pfam; PF02867; Ribonuc\_red\_1c; 1.  
 DR PRINTS; PR01183; RIBORDTASEM1.  
 KW Oxidoreductase.  
 FT NON\_TER 1 173  
 FT MON\_TER 173  
 SQ SEQUENCE 173 AA; 19697 MW; 33C10E3B458311A CRC64;

Query Match 42.9%; Score 6; DB 2; Length 173;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 IRKSKI 7  
 Db 43 IRKSKI 48

## RESULT 39

OG6H08 PRELIMINARY; PRT; 174 AA.  
 AC OG6H08;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=BT9727.2492;  
 OS Bacillus thuringiensis serovar konkukian str. 97-27.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;  
 OC Bacillus thuringiensis serovar konkukian.  
 OX NCBI\_TaxID=281309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=97-27;  
 RA Bretin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
 RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,  
 RA Richardson P., Rubin E., Tice H.;  
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB017355; AAT61854.1; -  
 DR InterPro; IPR011235; DUF\_UCP032285.  
 DR PIRSF; PIRSF032285; UCP032285; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 174 AA; 20427 MW; 43AFABF723FCB2 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 174;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RKSKIT 8  
 Db 57 RKSKIT 62

## RESULT 40

OG8IP59 PRELIMINARY; PRT; 174 AA.  
 AC OG8IP59; OG8K26; OG8K15;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocusNames=BA2718, BAS2532; ORFNames=GMAA2718.  
 OS Bacillus anthracis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1392;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate Porton;  
 RX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;  
 RA Read T.D., Peterson S.N., Tourasse N.J., Baillye L.W., Paulsen I.T.,  
 RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
 RA Holtzapple E.K., Ostred O.A., Helgason E., Ristone J., Wu M.,  
 RA Kolonay J.F., Beaman M.J., Dodson R.J., Brinkac L.M., Gilm M.L.,  
 RA Deboy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
 RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,  
 RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,  
 RA Berry K.J., Plaut R.D., Wolf A.M., Watkins J.L., Nierman W.C.,  
 RA Hazen A., Cline R.T., Redmond C., Thwaitte J.E., White O.,  
 RA Salzberg S.L., Thomson B., Friedlander A.M., Koehler T.M.,  
 RA Hanna P.C., Kolstoe A.B., Fraser C.M.;  
 RT "The genome sequence of Bacillus anthracis Ames and comparison to  
 closely related bacteria."  
 RL Nature 423:81-86 (2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate 0581;  
 RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,  
 RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,  
 RA Fraser C.M.;  
 RT "Bacillus anthracis comparative genomics."  
 RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.

```
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Serne; Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Brettin T.S., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Hitchcock P., Rubin E., Tice H.;
RA Richardson P., Rubin E., Tice H.;
RL Submitted (JAN-2004) to the EMBL/Genbank/DBJ databases.
DR EMBL; AE017032; AAP26558.1; -
DR EMBL; AE017334; AAT31833.1; -
DR EMBL; AE017225; AAT54843.1; -
DR TIGR; BA2718; -
DR InterPro; IPR011235; DUF UCP032285.
DR PIRSF; PIRSF032285; UCP032285; 1.
KW Hypothetical protein.
SQ SEQUENCE 174 AA; 20425 MW; BA90063C0DC85153 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 RSKIT 8
Db 57 RSKIT 62
```

Search completed: December 30, 2004, 15:37:17  
Job time : 199 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: December 30, 2004, 15:16:09 : Search time 38 Seconds  
(without alignments)  
35.448 Million cell updates/sec

Title: US-10-719-385-20  
Perfect score: 14  
Sequence: 1 MIRSKRTSVLSFC 14

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 150 summaries

Database : PIR 79:\*

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	50.0	1601	2 T18800	hypothetical prote
2	6	42.9	56	2 AH0071	hypothetical prote
3	6	42.9	75	2 T4139	hypothetical prote
4	6	42.9	143	1 R6MX15	ribosomal protein
5	6	42.9	159	2 T15627	hypothetical prote
6	6	42.9	184	1 TPBB1W	troponin I, slow 8
7	6	42.9	187	2 B44786	troponin I, slow 8
8	6	42.9	200	2 C49600	coat protein 22k -
9	6	42.9	255	2 T25853	hypothetical prote
10	6	42.9	316	2 D89830	hypothetical prote
11	6	42.9	324	2 S70390	DMC1/LIM5 homolog
12	6	42.9	326	2 T51811	protein farnesyltr
13	6	42.9	327	2 T00797	hypothetical prote
14	6	42.9	340	2 T49006	farnesyltransferas
15	6	42.9	343	2 A75597	mannosyltransferas
16	6	42.9	404	2 D95233	aminotransferase,
17	6	42.9	404	2 P98097	aspartate transfer
18	6	42.9	429	2 P86240	hypothetical prote
19	6	42.9	433	1 D8ECS8	homoserine dehydro
20	6	42.9	433	2 A83728	adenylosuccinate 1
21	6	42.9	469	2 D84949	NADH2 dehydrogenas
22	6	42.9	482	2 G86227	hypothetical prote
23	6	42.9	543	2 S45492	isp5 protein - fis
24	6	42.9	560	2 C84632	hypothetical prote
25	6	42.9	566	2 S19063	hypothetical prote
26	6	42.9	572	2 AF0211	probable ABC trans
27	6	42.9	580	2 T50059	sexual differentiation
28	6	42.9	587	2 S58319	hypothetical prote
29	6	42.9	644	2 D85359	hypothetical prote

30	6	42.9	651	2 G64068	DNA topoisomerase
31	6	42.9	652	2 H86221	hypothetical prote
32	6	42.9	693	2 AC0323	ribonucleoside-dip
33	6	42.9	727	2 T23585	hypothetical prote
34	6	42.9	809	2 C96552	hypothetical prote
35	6	42.9	809	2 T41645	probable spindle p
36	6	42.9	840	2 T33217	hypothetical prote
37	6	42.9	862	2 T49583	differentiation an
38	6	42.9	909	2 G69599	aconitate hydratase
39	6	42.9	1205	2 T27053	hypothetical prote
40	6	42.9	1242	2 S78061	DNA-directed RNA p
41	6	42.9	1377	2 D90538	hypothetical prote
42	6	42.9	1385	2 H88569	protein K03H1.5 [i
43	6	42.9	1409	2 S41028	hypothetical prote
44	6	42.9	1603	2 D89407	protein R10B8.6 [i
45	6	42.9	1696	2 T24146	hypothetical prote
46	5	35.7	53	2 S41957	narlingenin-chalcon
47	5	35.7	56	2 S45027	sox1 protein - Her
48	5	35.7	57	1 W9BPC7	gene 19.3 protein
49	5	35.7	59	2 G91095	hypothetical prote
50	5	35.7	59	2 C85941	hypothetical prote
51	5	35.7	60	2 D97849	hypothetical prote
52	5	35.7	70	2 T17923	hypothetical prote
53	5	35.7	72	1 QHEC1	heat-stable entero
54	5	35.7	72	1 QHEC4	heat-stable entero
55	5	35.7	72	1 QHEC3	heat-stable entero
56	5	35.7	80	2 E90686	hypothetical prote
57	5	35.7	80	2 D69745	hypothetical prote
58	5	35.7	80	2 A85537	unknown [imported]
59	5	35.7	83	2 H71068	hypothetical prote
60	5	35.7	87	2 AF0910	conserved hypochet
61	5	35.7	92	2 E95163	conserved hypochet
62	5	35.7	92	2 D98029	conserved hypochet
63	5	35.7	96	2 F64319	hypothetical prote
64	5	35.7	98	2 B97821	hypothetical prote
65	5	35.7	98	2 A97632	hypothetical prote
66	5	35.7	99	2 AD3031	hypothetical prote
67	5	35.7	100	1 S26080	ribosomal protein
68	5	35.7	103	2 C75149	hypothetical prote
69	5	35.7	104	2 H65115	hypothetical prote
70	5	35.7	104	2 B85968	hypothetical prote
71	5	35.7	104	2 G91142	hypothetical prote
72	5	35.7	105	2 F86774	hypothetical prote
73	5	35.7	106	2 A69027	ribosomal protein
74	5	35.7	107	2 A49442	Ig heavy chain V r
75	5	35.7	107	2 S61135	hypothetical prote
76	5	35.7	107	2 T51124	Iycopene cyclase (
77	5	35.7	109	1 PVRYC	parvalbumin - thox
78	5	35.7	113	2 S26465	Ig heavy chain V r
79	5	35.7	114	2 B83798	hypothetical prote
80	5	35.7	115	2 D64710	hypothetical prote
81	5	35.7	115	2 A71809	hypothetical prote
82	5	35.7	116	2 S68195	probable membrane
83	5	35.7	117	2 C69969	hypothetical prote
84	5	35.7	118	1 B64248	hypothetical prote
85	5	35.7	122	2 S11740	Ig heavy chain pre
86	5	35.7	122	2 S69460	hypothetical prote
87	5	35.7	125	2 JU0179	heparin-binding pr
88	5	35.7	126	2 JC7571	fatty acid-binding
89	5	35.7	128	1 S52084	ribosomal protein
90	5	35.7	128	2 S33611	narlingenin-chalcon
91	5	35.7	128	2 JC2130	heparin-binding pr
92	5	35.7	128	2 JC2121	heparin-binding pr
93	5	35.7	128	2 JC2119	heparin-binding pr
94	5	35.7	130	2 H86832	hypothetical prote
95	5	35.7	132	1 T43826	ribosomal protein
96	5	35.7	132	2 F84332	50S ribosomal prot
97	5	35.7	132	2 B84543	actin depolymerizi
98	5	35.7	136	2 A69461	virulence associat
99	5	35.7	136	2 B90058	conserved hypochet
100	5	35.7	137	2 A11176	E. coli phm prote
101	5	35.7	142	2 E90155	hypothetical prote
102	5	35.7	142	2 JQ1865	hypothetical 15.1K

103	5	35.7	143	2	PT0174	Ig heavy chain pre
104	5	35.7	143	2	JN0093	hypothetical 17k p
105	5	35.7	143	2	A71050	hypothetical prote
106	5	35.7	145	1	U00347	capsid protein VP3
107	5	35.7	145	2	G69011	hypothetical prote
108	5	35.7	145	2	AD3143	conserved hypotet
109	5	35.7	145	2	H98144	hypothetical prote
110	5	35.7	146	2	AE0982	probable acetyltra
111	5	35.7	147	2	AB4546	50S ribosomal prot
112	5	35.7	148	2	AG1645	hypothetical prote
113	5	35.7	148	2	AE1594	hypothetical prote
114	5	35.7	150	2	G75504	hypothetical prote
115	5	35.7	152	2	T07858	glycine-rich prote
116	5	35.7	157	2	G82080	hypothetical prote
117	5	35.7	160	2	A70815	probable mos prote
118	5	35.7	162	2	T49957	ribosomal protein
119	5	35.7	166	2	G86167	protein F2187.33 l
120	5	35.7	167	2	T00888	hypothetical prote
121	5	35.7	169	2	AE1166	transcription regu
122	5	35.7	170	2	T11964	allophycocyanin be
123	5	35.7	171	2	D66444	probable heat shoc
124	5	35.7	174	2	G70220	exported protein A
125	5	35.7	175	2	G72683	hypothetical prote
126	5	35.7	175	2	G96532	hypothetical prote
127	5	35.7	175	2	T28873	hypothetical prote
128	5	35.7	176	2	H70196	adenine phosphorib
129	5	35.7	177	2	T25232	hypothetical prote
130	5	35.7	177	2	H80299	hypothetical prote
131	5	35.7	179	2	B90408	hypothetical prote
132	5	35.7	180	2	B96619	protein T30316.5 l
133	5	35.7	181	2	B90588	hypothetical prote
134	5	35.7	181	2	AF1913	hypothetical prote
135	5	35.7	181	2	B69844	lytic transglycosy
136	5	35.7	181	2	T28270	ORF MSV110 hypote
137	5	35.7	182	2	H80288	conserved hypotet
138	5	35.7	187	1	E70443	L-fuculose-phospha
139	5	35.7	191	2	S69735	hypothetical prote
140	5	35.7	197	2	JC7735	frizzled-related p
141	5	35.7	202	2	F97342	transcription regu
142	5	35.7	203	2	T20111	hypothetical prote
143	5	35.7	203	2	A43856	major fibrinolytic
144	5	35.7	206	2	G76189	probable marf-fam1
145	5	35.7	209	2	C97461	hypothetical prote
146	5	35.7	212	2	B89891	hypothetical prote
147	5	35.7	213	2	S30181	ABC-transporting t
148	5	35.7	213	2	AB3685	nascent polypeptid
149	5	35.7	215	2	S49326	nascent polypeptid
150	5	35.7	215	2	T30827	nascent polypeptid

## ALIGNMENTS

## RESULT 1

T18800 hypothetical protein C01B9.1 - *Caenorhabditis elegans*

C/Species: *Caenorhabditis elegans*

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C/Accession: T18800

R/Query: C.

submitted to the EMBL Data Library, March 1997

A/Reference number: Z19024

A/Accession: T18800

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1601 <WIL>

A/Cross-references: UNIPROT:O45231, EMBL:Z93373, P1DN:CA807551.1, GSPDB:GN00020, CESP:CD

A/Experimental source: clone C01B9

C/Genes: CESP:C01B9.1

A/Map position: 2

A/Intons: 151/3; 243/2; 478/3; 569/2; 795/3; 879/3; 895/2; 1227/3; 1270/2; 1514/3

Query Match	50.0%;	Score 7;	DB 2;	Length 1601;
Best Local Similarity	100.0%;	Pred. No. 21;		
Matches	7;	Conservative 0;	Mismatches 0;	Indels 0;

QY	7	ITSVLSF 13
DB	398	ITSVLSF 404

## RESULT 2

AH0071 hypothetical protein YP00578 [imported] - *Yersinia pestis* (strain CO92)

C/Species: *Yersinia pestis*

C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004

C/Accession: AH0071

R/Query: J.; Wren, B.W.; Thomson, N.R.; Tilball, R.W.; Holden, M.T.G.; Prentice, M.B.;

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, J.

Nature 413, 523-527, 2001

A/Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AH0071

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-56 <KUR>

A/Cross-references: UNIPROT:Q8ZIC7; GB:AL590842; P1DN:CA89435.1; P1D:G15978671; GSPDB:G

C/Genes: YP00578

Query Match	42.9%;	Score 6;	DB 2;	Length 56;
Best Local Similarity	100.0%;	Pred. No. 14;		
Matches	6;	Conservative 0;	Mismatches 0;	Indels 0;

QY	7	ITSVLS 12
DB	48	ITSVLS 53

## RESULT 3

T44139 hypothetical protein [imported] - *Staphylococcus aureus* (fragment)

C/Species: *Staphylococcus aureus*

C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 11-May-2000

C/Accession: T44139

R/Query: T.; Katayama, Y.; Hiramatsu, K.

Antimicrob. Agents Chemother. 43, 1449-1458, 1999

A/Title: Cloning and nucleotide sequence determination of the entire mec DNA of pre-meth

A/Reference number: Z22733; MUID:99278010; PMID:10348769

A/Accession: T44139

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-75 <ITO>

A/Cross-references: EMBL:D86934; P1DN:BA82241.1

A/Experimental source: strain N315

Query Match	42.9%;	Score 6;	DB 2;	Length 75;
Best Local Similarity	100.0%;	Pred. No. 18;		
Matches	6;	Conservative 0;	Mismatches 0;	Indels 0;

QY	9	SVLSFC 14
DB	48	SVLSFC 53

## RESULT 4

R6X15 ribosomal protein L15 - *Methanococcus vannielii*

C/Species: *Methanococcus vannielii*

C/Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 09-Jul-2004

C/Accession: S05626

R/Query: J.; Spicker, G.; Boeck, A.

U. Mol. Biol. 209, 21-36, 1989

A/Title: Organization and structure of the *Methanococcus* transcriptional unit homologous

S. ribosomes.  
 A:Reference number: 505611; MUID:90040717; PMID:2530355  
 A:Accession: S05626  
 A:Molecule type: DNA  
 A:Residues: 1-143 <AUE>  
 A:Cross-references: UNIPROT:P14032; EMBL:X16270; NID:G44754; PIDN:CAA34702.1; PID:G44770  
 C:Superfamily: rat ribosomal protein L27a  
 C:Keywords: protein biosynthesis; ribosome

Query Match 42.9%; Score 6; DB 1; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRRSK 6  
 Db 1 MIRRSK 6

RESULT 5  
 T15627  
 hypothetical protein C25H3.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jan-2000  
 C:Accession: T15627  
 R:Johnson, D.  
 submitted to the EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid C25H3.  
 A:Reference number: Z18379  
 A:Accession: T15627  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Residues: 1-159 <JOH>  
 A:Cross-references: EMBL:U29535; NID:G860251; PIDN:AAA68784.1; CESP:C25H3.5  
 A:Experimental source: strain Bristol N2  
 C:Genetics:  
 A:Gene: CESP:C25H3.5  
 A:Insertions: 21/1  
 C:Superfamily: Caenorhabditis elegans hypothetical protein C25H3.5

Query Match 42.9%; Score 6; DB 2; Length 159;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRRSK 6  
 Db 106 MIRRSK 111

RESULT 6  
 TPRB1W  
 tropomyosin I, slow skeletal muscle - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 30-Apr-1979 #sequence\_revision 30-Apr-1979 #text\_change 09-Jul-2004  
 C:Accession: A03089  
 R:Grand, R.J.A.; Wilkinson, J.M.  
 Biochem. J. 167, 183-192, 1977

A:Title: The amino acid sequence of rabbit slow-muscle tropomyosin I.  
 A:Reference number: A90296; MUID:78060292; PMID:588250  
 A:Accession: A03089  
 A:Molecule type: protein  
 A:Residues: 1-184 <GRA>

A:Cross-references: UNIPROT:P02645  
 A>Note: some of the molecules lack residues 183 and 184  
 C:Complex: tropomyosin is a heterotrimer with one molecule each of tropomyosin C (calcium bind  
 C:Function:  
 A:Description: binds actin and inhibits myosin ATPase activity; with tropomyosin mediate

A:Pathway: muscle contraction  
 C:Superfamily: tropomyosin I  
 C:Keywords: actin binding; blocked amino end  
 F:1/Modified site: blocked amino end (Pro) (partial) (probably acetylated) #status exper

Query Match 42.9%; Score 6; DB 1; Length 184;  
 Best Local Similarity 100.0%; Pred. No. 37;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RKSKIT 8  
 Db 5 RKSKIT 10

RESULT 7  
 B44786

tropomyosin I, slow skeletal muscle - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 19-Mar-1993 #sequence\_revision 19-Mar-1993 #text\_change 09-Jul-2004  
 C:Accession: B44786  
 R:Koppe, R.I.; Hallauer, P.L.; Karpaci, G.; Haefliger, K.E.M.  
 J. Biol. Chem. 264, 14327-14333, 1989

A:Title: cDNA clone and expression analysis of rodent fast and slow skeletal muscle tropo  
 A:Reference number: A44786; MUID:89340548; PMID:2760067  
 A:Accession: B44786  
 A:Status: preliminary  
 A:Molecule type: mRNA

A:Residues: 1-187 <KOP>  
 A:Cross-references: UNIPROT:P13413; GB:U04993; NID:G207517; PIDN:AAA42295.1; PID:G207518  
 C:Superfamily: tropomyosin I  
 C:Keywords: skeletal muscle

Query Match 42.9%; Score 6; DB 2; Length 187;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RKSKIT 8  
 Db 6 RKSKIT 11

RESULT 8  
 C49600

coat protein 22k - soybean dwarf virus

C:Species: soybean dwarf virus  
 C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004  
 C:Accession: C49600  
 R:Rathjen, J.P.; Karsgeorgos, L.E.; Habili, N.; Waterhouse, P.M.; Symons, R.H.  
 Virology 198, 671-679, 1994

A:Title: Soybean dwarf luteovirus contains the third variant genome type in the luteovir  
 A:Reference number: A49600; MUID:94120742; PMID:8251248  
 A:Accession: C49600

A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-200 <RAT>  
 A:Cross-references: UNIPROT:O87033; GB:U24049; NID:G436017; PIDN:AAA17537.1; PID:G436020  
 C:Superfamily: potato leaf roll virus coat protein; potato leaf roll virus coat protein  
 F:1-200/Domain: potato leaf roll virus coat protein homology <COP>

Query Match 42.9%; Score 6; DB 2; Length 200;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KITSVL 11  
 Db 104 KITSVL 109

RESULT 9  
 T25853

hypothetical protein T01B11.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C:Accession: T25853  
 R:Geisel, C.; Stellyes, L.  
 submitted to the EMBL Data Library, December 1996

A:Description: The sequence of C. elegans cosmid T01B11.  
 A:Reference number: Z20099  
 A:Accession: T25853  
 A:Status: preliminary; translated from GB/EMBL/DBJ



A:Molecule type: DNA  
A:Residues: 1-255 <GE1>  
A:Cross-references: UNIPROT:Q9N9F1, EMBL:U80931, PIDDN:AA838004.1, GSPDB:GN00022, CESP:TD  
A:Experimental source: Strain Bristol N2; clone T01B11  
C:Genetics:  
A:Gene: CESP:T01B11.7  
A:Map position: 4  
A:introns: 72/1; 114/3; 154/3; 195/2; 214/3

Query Match 42.9%; Score 6; DB 2; Length 255;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TSUSF 13  
Db 246 TSUSF 251

RESULT 10  
D89830  
hypothetical protein SA0567 [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: D89830  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguma, A.; Mizutani, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratake, K.  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: D89830  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-316 <KUR>  
A:Cross-references: UNIPROT:Q9W02; GB:BA000018; PID:g13700502; PIDDN:BB41799.1; GSPDB:C  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA0567  
C:Superfamily: vitamin B12 transport protein btuc

Query Match 42.9%; Score 6; DB 2; Length 316;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ITSVLS 12  
Db 118 ITSVLS 123

RESULT 11  
S70390  
DMC1/LIM15 homolog 1 - yeast (Candida albicans)  
C:Species: Candida albicans  
C>Date: 06-Dec-1996 #sequence\_revision 13-Mar-1997 #text\_change 09-Jul-2004  
C:Accession: S70390  
R:Diener, A.C.; Fink, G.R.  
A:Title: DMC1 is a functional Candida albicans homologue of the meiosis-specific gene DMC1.  
A:Reference number: S70390; MUID:96363911; PMID:8725225  
A:Accession: S70390  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-324 <DIR>  
A:Cross-references: UNIPROT:P50265; EMBL:U39808; NID:g1145715; PIDDN:AAC49400.1; PID:g114  
C:Genetics:  
A:Gene: DMC1  
A:introns: 35/3; 67/3  
C:Superfamily: yeast DNA repair protein RAD51

Query Match 42.9%; Score 6; DB 2; Length 324;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ITSVLS 12  
Db 34 ITSVLS 39

RESULT 12  
T5181  
protein farnesyltransferase (BC 2.5.1.-) chain A [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 09-Jul-2004  
C:Accession: T5181  
R:Belbhir, L.; Villarejo, R.; Inze, D.; Thomas, D.; Thomasset, B.  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: Z25461  
A:Accession: T5181  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-326 <BEL>  
A:Cross-references: UNIPROT:Q9LX33; EMBL:AF064542; PIDDN:AAC61853.1  
C:Genetics:  
A:Gene: FTA  
C:Keywords: transferase

Query Match 42.9%; Score 6; DB 2; Length 326;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RKSKIT 8  
Db .317 RKSKIT 322

RESULT 13  
T00797  
hypothetical protein At2g32710 [imported] - Arabidopsis thaliana  
N:Alternate names: hypothetical protein F24L7.15  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
C:Accession: T00797; E84736  
R:Rounsley, S.D.; Lin, X.; Kechum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, R.  
A:Title: The Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.  
A:Reference number: Z14204  
A:Accession: T00797  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-327 <ROU>  
A:Cross-references: UNIPROT:O48846; EMBL:AC003974; NID:g2914688; PID:g2914702  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentto, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euse, D.; Niemman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: E84736  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-327 <STO>  
A:Cross-references: GB:AE002093; NID:g2914702; PIDDN:AAC04492.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g32710; F24L7.15  
A:Map position: 2  
A:introns: 192/2  
C:Superfamily: Arabidopsis thaliana hypothetical protein F24L7.15

Query Match 42.9%; Score 6; DB 2; Length 327;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 IRKSKI 7  
Db 5 IRKSKI 10

## RESULT 14

T49006  
A:Accession: T49006  
A:Title: Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-340 <DNA>  
A:Cross-references: UNIPROT:O9LX33; EMBL:AL356014; GSPDB:GN00061; ATSP:F25L23.240  
A:Experimental source: cultivar Columbia; BAC clone F25L23  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 09-Jul-2004  
C:Accession: T49006  
R:ID Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; L submitted to the Protein Sequence Database, May 2000  
A:Accession: T49006  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-340 <DNA>  
A:Cross-references: UNIPROT:O9LX33; EMBL:AL356014; GSPDB:GN00061; ATSP:F25L23.240  
C:Genetics:  
A:Gene: ATSP:F25L23.240  
A:Map position: 3  
A:introns: 89/3; 126/2; 166/3; 198/3

Query Match 42.9%; Score 6; DB 2; Length 340;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSKITS 8  
Db 331 RSKITS 336

## RESULT 15

A75597  
A:Accession: A75597  
A:Title: Complete genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Cross-references: UNIPROT:O9R2B5; GB:AE001662; GB:AE001825; NID:g6460468; PIDN:AAF1227  
C:Genetics:  
A:Gene: DRA0039  
A:Map position: 2  
Query Match 42.9%; Score 6; DB 2; Length 343;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVS 12  
Db 24 ITSVS 29

## RESULT 16

D95233  
A:Accession: D95233  
A:Title: Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-404 <DNA>  
A:Cross-references: UNIPROT:O97NM6; GB:AE005672; PIDN:AAK76061.1; PID:g14973502; GSPDB:G  
A:Experimental source: strain TIGR4  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: D95233  
R:Teitelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-404 <DNA>  
A:Cross-references: UNIPROT:O97NM6; GB:AE005672; PIDN:AAK76061.1; PID:g14973502; GSPDB:G  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP1994  
C:Superfamily: aspartate transaminase

Query Match 42.9%; Score 6; DB 2; Length 404;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKITS 9  
Db 162 KSKITS 167

## RESULT 17

F98097  
A:Accession: F98097  
A:Title: Complete genome sequence of the bacterium *Streptococcus pneumoniae* Strain R6.  
A:Reference number: A97872; MUID:2143245; PMID:11544234  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-404 <DNA>  
A:Cross-references: UNIPROT:O8DNC7; GB:AE007317; PIDN:AAU0611.1; PID:g15459494; GSPDB:G  
C:Genetics:  
A:Gene: aspc  
C:Superfamily: aspartate transaminase  
C:Keywords: aminotransferase  
Query Match 42.9%; Score 6; DB 2; Length 404;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKITS 9  
Db 162 KSKITS 167

## RESULT 18

F86240  
A:Accession: F86240  
A:Title: Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-404 <DNA>  
A:Cross-references: UNIPROT:O97NM6; GB:AE005672; PIDN:AAK76061.1; PID:g14973502; GSPDB:G  
A:Experimental source: strain TIGR4  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: F86240  
R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alonso, Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hultar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A:Reference number: A66141; MUID:21016719; PMID:11130712  
A:Accession: F86240  
A:Status: Preliminary

A:Molecule type: DNA  
 A:Residues: 1-429 <STO>  
 A:Cross-references: UNIPROT:Q9SAD0; GB:AE005172; NID:g4874276; PIND:AA031341.1; GSPDB:GN  
 C:Genetics:  
 A:Map position: 1

Query Match 42.9%; Score 6; DB 2; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 SKITS 9  
 DB 416 SKITS 421

## RESULT 19

DESCRIPTION  
 homoserine dehydrogenase (EC 1.1.1.3) - *Bacillus subtilis*

C:Species: *Bacillus subtilis*

C>Date: 30-Jun-1991 #sequence\_revision 02-Jul-1998 #text\_change 09-Jul-2004

C:Accession: A31973; C25364; D69642

R:Parrot, C.; Cohen, G.N.

J. Biol. Chem. 263, 14654-14660, 1988

A:Title: Cloning and nucleotide sequence of the *Bacillus subtilis* hom gene coding for ho  
 ne dehydrogenases I and II.

A:Reference number: A31973; MUID:89008330; PMID:3139660

A:Accession: A31973

A:Molecule type: DNA

A:Residues: 1-401, 'O', 403-433 <PAR1>

A:Cross-references: UNIPROT:P1982; EMBL:M23217; NID:g340846; PIND:AAA50609.1; PID:g5584

A:Note: These authors used TTG as a start codon; the codon given for 402-Glu (CAA) is IT

R:Parrot, C.

EMBO J. 5, 3013-3019, 1986

A:Title: Evolution of biosynthetic pathways: a common ancestor for threonine synthase, t

A:Reference number: A91055; MUID:87080286; PMID:3098560

A:Accession: C25364

A:Molecule type: DNA

A:Residues: 353-374, 'T', 376-433 <PAR2>

A:Cross-references: GB:X04603; NID:g40210; PIND:CA28269.1; PID:8809663

R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bester

A:; Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carer, N.M.; Cho

A:; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallier

A:; J. J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

A:; Koetter, P.; Koningsstein, G.; Krogsh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A:; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel

A:; Y. M.; Ogawa, K.; Ogiwara, A.; Oudaga, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portecelle

A:; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,

A:; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowaka, A.; Serot

A:; akouchi, M.; Tamakoshi, A.; Tanaka, T.; Tarpetira, P.; Tognoni, A.; Tosato, V.; Uchiyama,

A:; T.; Winters, P.; Wpat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:; Authors: Yoshikawa, H.F.; Zumestein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: D69642

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-433 <KUN>

A:Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; PIND:CAB15216.1; PID:g2635723

A:Experimental source: strain 168

C:Genetics:

A:Gene: hom

A:Map position: 78 min

A:Function:  
 A:Description: catalyzes the reduction of L-aspartate 4-semialdehyde to L-homoserine by

C:Superfamily: homoserine dehydrogenase; homoserine dehydrogenase homology

C:Keywords: NADP, oxidoreductase; threonine biosynthesis

F:1-243/Domain: homoserine dehydrogenase homology <HSD>

F:5-33/Region: beta-alpha-beta NADP nucleotide-binding fold

Query Match 42.9%; Score 6; DB 1; Length 433;

Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
 DB 363 SKITSV 368

## RESULT 20

A83728

adenylosuccinate lyase purB [imported] - *Bacillus halodurans* (strain C-125)

C:Species: *Bacillus halodurans*

C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004

C:Accession: A83728

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: A83728

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-433 <STO>

A:Cross-references: UNIPROT:Q9K61; GB:AB001509; GB:BA000004; NID:g10173176; PIND:BA043

A:Experimental source: strain C-125

C:Genetics:

A:Gene: purB

C:Superfamily: fumarate hydratase

Query Match 42.9%; Score 6; DB 2; Length 433;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
 DB 399 ITSVLS 404

## RESULT 21

D84949

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain N [imported] - *Buchnera* sp. (strain

C:Species: *Buchnera* sp.

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 03-Jun-2002

C:Accession: D84949

R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

Nature 407, 81-86, 2000

A:Title: Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. A

A:Reference number: A84930; MUID:20445173; PMID:10993077

A:Accession: D84949

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-469 <STO>

A:Cross-references: GB:AP000398; GSPDB:GN00144

A:Experimental source: strain APS

C:Genetics:

A:Gene: nuon; BUI66

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 2

C:Keywords: NAD, oxidoreductase

Query Match 42.9%; Score 6; DB 2; Length 469;  
 Best Local Similarity 100.0%; Pred. No. 80;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSVLSF 13  
 DB 220 TSVLSF 225

## RESULT 22

G86227

hypothetical protein [imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C:Accession: G86227

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.M.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart, K.; Jensen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Maiti, R.; Mazziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A:Reference number: A66141; MUID:21016719; PMID:11130712  
A:Accession: G86227  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-482 <STO>  
A:Cross-references: UNIPROT:O80527; GB:AE005172; NID:g3482919; PIDN:AAC33204.1; GSPDB:GN C:Genetics:  
A:Map position: 1

Query Match 42.9%; Score 6; DB 2; Length 482;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSFC 14  
Db 23 SVLSFC 28

RESULT 23  
S45492  
1aps protein - fission yeast (*Schizosaccharomyces pombe*)  
C:Species: *Schizosaccharomyces pombe*  
C:Date: 10-Dec-1994 #sequence\_revision 21-Jul-1995 #text\_change 09-Jul-2004  
A:Accession: S45492; S35896  
R:Sato, S.; Suzuki, H.; Widyastuti, U.; Hotta, Y.; Tabata, S.  
Curr. Genet. 26, 31-37, 1994  
A:Title: Identification and characterization of genes induced during sexual differentiation  
A:Reference number: S45492; MUID:95042833; PMID:7954893  
A:Accession: S45492  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-543 <SAT>  
A:Cross-references: UNIPROT:P40901; EMBL:D14062; NID:g218544; PIDN:BA03148.1; PID:g2185  
C:Genetics:  
A:Gene: 1aps  
C:Superfamily: arginine permease  
C:Keywords: transmembrane protein

Query Match 42.9%; Score 6; DB 2; Length 543;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSFC 14  
Db 537 SVLSFC 542

RESULT 24  
C84632  
hypothetical protein At2g24070 [imported] - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
A:Accession: C84632  
R:Lin, X.; Kaul, S.; Rounsa, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
A:Reference number: A64420; MUID:20083487; PMID:10617197  
A:Accession: C84632  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-560 <STO>

A:Cross-references: UNIPROT:O82235; GB:AE002093; NID:g3738334; PIDN:AAC63675.1; GSPDB:GN C:Genetics:  
A:Gene: At2g24070  
A:Map position: 2

Query Match 42.9%; Score 6; DB 2; Length 560;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSVSF 13  
Db 365 TSVSF 370

RESULT 25  
S19063  
hypothetical protein YNR023w precursor, mitochondrial - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: hypothetical protein N3224  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 09-Jul-2004  
A:Accession: S19063; S63354  
R:Donald, K.A.G.; Hill, J.; Griffiths, D.E.  
submitted to the EMBL Data Library, October 1991  
A:Reference number: S19063  
A:Accession: S19063  
A:Molecule type: DNA  
A:Residues: 1-566 <DON>  
A:Cross-references: UNIPROT:P53628; EMBL:X62430; NID:g4106; PID:g4107  
R:Pohl, T.M.  
submitted to the Protein Sequence Database, April 1996  
A:Reference number: S63346  
A:Accession: S63354  
A:Molecule type: DNA  
A:Residues: 1-566 <POH>  
A:Cross-references: EMBL:Z71638; NID:g1302511; PID:e239821; PID:g1302512; MIPS:YNR023w  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:SNF12  
A:Cross-references: SGD:S0005306; MIPS:YNR023w  
A:Map position: 14R  
A:Genome: nuclear  
C:Keywords: mitochondrion

Query Match 42.9%; Score 6; DB 2; Length 566;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
Db 487 ITSVLS 492

RESULT 26  
AF0211  
probable ABC transporter (ATP-binding protein) YPO1735 [imported] - *Yersinia pestis* (str  
C:Species: *Yersinia pestis*  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
A:Accession: AF0211  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tlball, R.W.; Holden, M.T.G.; Prentice, M.B. deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; li, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell, Nature 413, 523-527, 2001  
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AF0211  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-572 <KUP>  
A:Cross-references: UNIPROT:O8ZFR4; GB:AL590842; PIDN:CAC90554.1; PID:g15979763; GSPDB:G C:Genetics:  
A:Gene: YPO1735  
C:Superfamily: Escherichia coli ABC transporter mdla; ATP-binding cassette homology

Query Match 42.9%; Score 6; DB 2; Length 572;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 6 KITSVL 11  
Db 289 KITSVL 294

RESULT 27  
T50059  
sexual differentiation process probable amino-acid permease isps [imported] - fission ye  
C:Species: Schizosaccharomyces pombe  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jul-2004  
C:Accession: T50059  
R: Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, December 1999  
A:Reference number: Z25031  
A:Accession: T50059  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-580 <HUN>  
A:Cross-references: UNIPROT:P40901; EMBL:AL133521; PIDN:CAB63545.1; GSPDB:GN00066; SPDB:  
C:Genetics:  
A:Gene: SPDB:SPAC1039.09  
A:Map position: 1  
C:Superfamily: arginine permease

Query Match 42.9%; Score 6; DB 2; Length 580;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 9 SVLSFC 14  
Db 574 SVLSFC 579

RESULT 28  
S58319  
hypotheical protein YOR308C - Yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein O567  
C:Species: Saccharomyces cerevisiae  
C:Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 09-Jul-2004  
C:Accession: S58319; S67214; S71987  
R: Pearson, B.M.; Hernando, Y.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: S58318  
A:Accession: S58319  
A:Molecule type: DNA  
A:Residues: 1-587 <PEA>  
A:Cross-references: UNIPROT:Q12420; EMBL:X90565; NID:g940836; PID:g940838  
R: Pearson, B.M.; Hernando, Y.; Kalogeropoulos, A.; Schweizer, M.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67213  
A:Accession: S67214  
A:Molecule type: DNA  
A:Residues: 1-587 <PEA>  
A:Cross-references: EMBL:Z75216; NID:g1420678; PID:g252142; PID:g1420679; MIPS:YOR308C  
R: Czapluch, C.; Jantiaux, J.C.; Korde, E.; Poirey, R.; Pujol, A.; Tobiasch, E.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67194  
A:Accession: S67212  
A:Molecule type: DNA  
A:Residues: 285-587 <CZI>  
A:Cross-references: EMBL:Z75216; MIPS:YOR308C  
A:Experimental source: strain S288C  
R: Pearson, B.M.; Hernando, Y.; Payne, J.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.  
Yeast 12, 1021-1031, 1996  
A:Title: Sequencing of a 35.71 kb DNA segment on the right arm of yeast chromosome XV re  
A:Reference number: S71986; MUID:97051589; PMID:8896266  
A:Accession: S71987

A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-587 <PEA>  
A:Cross-references: EMBL:X90565; NID:g940836; PIDN:CAB62162.1; PID:g940838  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995  
C:Genetics:  
A:Gene: SGD:SNUG6  
A:Cross-references: SGD:S0005835  
A:Map position: 15R

Query Match 42.9%; Score 6; DB 2; Length 587;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 3 RKSKIT 8  
Db 294 RKSKIT 299

RESULT 29  
D85359  
hypotheical protein AT4g30710 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: D85359  
R: Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488; PMID:10617258  
A:Accession: D85359  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <STO>  
A:Cross-references: UNIPROT:Q9SUH5; GB:NC\_001268; NID:g726972; PIDN:CAB79789.1; GSPDB:G  
A:Gene: AT4g30710  
A:Map position: 4

Query Match 42.9%; Score 6; DB 2; Length 644;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 8 TSVLSF 13  
Db 410 TSVLSF 415

RESULT 30  
G64068  
DNA topoisomerase (EC 5.99.1.2) III - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 09-Jul-2004  
C:Accession: G64068  
R: Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: G64068  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-651 <TIGR>  
A:Cross-references: UNIPROT:P43704; GB:U32727; GB:U42023; NID:g1573415; PIDN:AAC22103.1;  
C:Superfamily: DNA topoisomerase I topa  
C:Keywords: DNA binding; isomerase

Query Match 42.9%; Score 6; DB 2; Length 651;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 3 RKSKIT 8

Db 431 RKSXIT 436

## RESULT 31

H86221

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C:Accession: H86221

A:Theologian: A.; Becker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart, K.;

ansen, N.F.; Hughes, B.; Huizlar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; PMID:21016719; PMID:11130712

A:Accession: H86221

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-652 &lt;STO&gt;

A:Cross-references: UNIPROT:Q04030; GB:AE005172; NID:G2342692; PIDN:AB70419.1; GSPDB:GN

C:Genetics:

A:Map position: 1

## Query Match

Best Local Similarity 42.9%; Score 6; DB 2; Length 652;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IRKSKI 7

Db 41 IRKSKI 46

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T23585

R:Steward, C.

submitted to the EMBL Data Library, March 1997

A:Reference number: 219764

A:Accession: T23585

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-727 <MIL>

A:Cross-references: UNIPROT:Q45675; EMBL:Z92806; PIDN:CA807259.1; GSPDB:GN00023; CESP:KI

A:Experimental source: clone K10G4

C:Genetics:

A:Gene: CESP:K10G4.4

A:Map position: 5

A:Introns: 168/3; 259/2; 494/3; 555/3

Query Match

Best Local Similarity 42.9%; Score 6; DB 2; Length 727;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSVLSP 13

Db 415 TSVLSP 420

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

A;Cross-references: UNIPROT:O94534, EMBL:AL035524, PDB:CAA22843.1, GSPDB:GN00068, SPDB  
A;Experimental source: strain 972h-, cosmid c895  
C;Genetic8:  
A;Gene: SPDB:SPCC895.07  
A;Map position: 3

Query Match	42.9%	Score 6	DB 2	Length 809
Best Local	Similarity 100.0%	Pred. No.	1.3e+02	
Matches 6	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	2	IRRSKI	7
Db	491	IRRSKI	496

RESULT 36  
T33217  
hypochemical protein T07H8.6 - *Caenorhabditis elegans*

Query Match	42.9%	Score 6	DB 2	Length 840
Best Local Similarity	100.0%	Pred. No.	1.3e+02	
Matches 6	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	8	TSVLSF	13
Db	375	TSVLSF	380

RESULT 37  
149583  
differentiation antigen - mouse

Query Match	42.9%	Score 6;	DB 2;	Length 862;
Best Local Similarity	100.0%	Pred. No. 1.3e+02;		
Matches 6;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	5	SKITSV	10
Db	185	SKITSV	190

### RESULT 38

G69599

Query Match	42.9%	Score 6;	DB 2;	Length 909;
Best Local Similarity	100.0%	Pred. No. 1.4e+02;		
Matches 6;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	7	ITSVLS	12
Db	541	ITSVLS	546

RESULT 39  
T27053  
hypothetical protein Y49E10.19 - *Caenorhabditis elegans*



A:Map position: 3  
A:Introns: 43/1; 281/3; 370/3; 418/3; 836/2; 949/3; 995/3; 1119/3

Query Match 42.9%; Score 6; DB 2; Length 1205;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SKITSV 10  
|||  
Db 34 SKITSV 39

## RESULT 40

S78061

DNA-directed RNA polymerase (EC 2.7.7.6) II - African swine fever virus

N/Alternate names: EPI242L protein

C/Species: African swine fever virus, ASFV

C/Date: 19-Nov-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004

C/Accession: S78061; S33996

R/Yanez, R.J.

submitted to the EMBL Data Library, January 1993

A/Reference number: S78060

A/Accession: S78061

A/Molecule type: DNA

A/Residues: 1-1242 <YAN>

A/Cross-references: UNIPROT:P42487; EMBL:Z21490; NID:g311527; PIDN:CAA79698.1; PID:g3115

R/Yanez, R.J.; Bourneil, M.; Nogai, M.L.; Yuste, L.; Vinuela, E.

Nucleic Acids Res. 21, 2423-2427, 1993

A/Title: African swine fever virus encodes two genes which share significant homology w

A/Reference number: S33995; MUID:93281390; PMID:8506138

A/Accession: S33996

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 99-212;367-428;538-679;770-880;910-1048;1074-1201 <YAN>

A/Cross-references: EMBL:Z21490

A/Experimental source: strain BA71V

C/Superfamily: DNA-directed RNA polymerase 132K polypeptide

C/Keywords: DNA binding; nucleic acid transferase; transcription; zinc finger

Query Match 42.9%; Score 6; DB 2; Length 1242;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ITSVLS 12  
|||  
Db 1230 ITSVLS 1235

Search completed: December 30, 2004, 15:33:59  
Job time : 45 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2004, 15:27:39 : Search time 37 Seconds

(without alignments)  
25.093 Million cell updates/sec

Title: US-10-719-385-20  
Perfect score: 14  
Sequence: 1 MIRSKITSVLSFC 14

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 150 summaries

Database : Issued Patents, AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	50.0	162	4	US-09-134-000C-3462
2	6	42.9	77	4	US-09-513-999C-4730
3	6	42.9	133	4	US-09-107-532A-4535
4	6	42.9	135	4	US-09-543-681A-4409
5	6	42.9	143	4	US-09-732-210-617
6	6	42.9	183	4	US-09-543-681A-7075
7	6	42.9	238	4	US-09-583-110-3590
8	6	42.9	270	4	US-09-107-532A-7734
9	6	42.9	275	4	US-09-248-796A-16584
10	6	42.9	311	4	US-09-489-039A-9786
11	6	42.9	327	4	US-09-526-597D-34
12	6	42.9	344	4	US-09-107-532A-5546
13	6	42.9	433	1	US-08-700-335-20
14	6	42.9	494	4	US-09-019-095A-24
15	6	42.9	540	4	US-08-945-771-2
16	6	42.9	661	4	US-09-371-338-7
17	6	42.9	739	3	US-09-134-001C-3586
18	6	42.9	967	3	US-09-543-681A-6407
19	5	35.7	12	3	US-09-177-249-124
20	5	35.7	13	3	US-09-522-666-30
21	5	35.7	15	1	US-08-268-221-5
22	5	35.7	15	1	PCR-US93-01112-5
23	5	35.7	16	3	US-09-360-237-55
24	5	35.7	16	4	US-09-556-877-251
25	5	35.7	16	4	US-09-620-412C-251
26	5	35.7	16	4	US-09-598-419-251
27	5	35.7	16	4	US-09-680-571A-83

28	5	35.7	16	4	US-09-680-571A-97	Sequence 97, Appl
29	5	35.7	20	4	US-09-556-877-248	Sequence 248, App
30	5	35.7	20	4	US-09-556-877-249	Sequence 249, App
31	5	35.7	20	4	US-09-556-877-250	Sequence 250, App
32	5	35.7	20	4	US-09-620-412C-248	Sequence 248, App
33	5	35.7	20	4	US-09-620-412C-249	Sequence 249, App
34	5	35.7	20	4	US-09-620-412C-250	Sequence 250, App
35	5	35.7	20	4	US-09-598-419-248	Sequence 248, App
36	5	35.7	20	4	US-09-598-419-249	Sequence 249, App
37	5	35.7	20	4	US-09-598-419-250	Sequence 250, App
38	5	35.7	20	4	US-09-834-759-545	Sequence 545, App
39	5	35.7	23	1	US-08-268-251-35	Sequence 35, Appl
40	5	35.7	23	1	PCR-US93-01112-35	Sequence 35, Appl
41	5	35.7	27	3	US-09-360-237-57	Sequence 57, Appl
42	5	35.7	32	4	US-09-759-112A-17	Sequence 17, Appl
43	5	35.7	37	4	US-09-207-388-17	Sequence 17, Appl
44	5	35.7	39	3	US-08-984-277-10	Sequence 10, Appl
45	5	35.7	40	3	US-09-286-691-17	Sequence 17, Appl
46	5	35.7	40	3	US-09-687-147-17	Sequence 17, Appl
47	5	35.7	44	4	US-09-443-780C-19	Sequence 19, Appl
48	5	35.7	44	4	US-09-079-723-3	Sequence 3, Appl
49	5	35.7	50	4	US-09-583-110-4798	Sequence 4798, App
50	5	35.7	51	4	US-09-621-976-7274	Sequence 7274, App
51	5	35.7	53	4	US-09-513-999C-7248	Sequence 7248, App
52	5	35.7	60	4	US-09-513-999C-4510	Sequence 4510, App
53	5	35.7	61	3	US-09-134-001C-4270	Sequence 4270, App
54	5	35.7	63	4	US-09-489-039A-7831	Sequence 7831, App
55	5	35.7	63	4	US-09-489-039A-10388	Sequence 10388, App
56	5	35.7	63	4	US-09-248-796A-25157	Sequence 25157, App
57	5	35.7	63	4	US-09-248-796A-25683	Sequence 25683, App
58	5	35.7	63	4	US-09-248-796A-27206	Sequence 27206, App
59	5	35.7	64	4	US-09-513-999C-4708	Sequence 4708, App
60	5	35.7	64	4	US-09-149-476-449	Sequence 449, App
61	5	35.7	64	4	US-09-270-767-54318	Sequence 54318, App
62	5	35.7	64	4	US-09-270-767-54635	Sequence 54635, App
63	5	35.7	65	4	US-09-328-352-5178	Sequence 5178, App
64	5	35.7	67	4	US-09-583-110-5275	Sequence 5275, App
65	5	35.7	68	4	US-09-583-110-4780	Sequence 4780, App
66	5	35.7	68	4	US-09-270-767-39522	Sequence 39522, App
67	5	35.7	68	4	US-09-270-767-54739	Sequence 54739, App
68	5	35.7	69	4	US-09-248-796A-41129	Sequence 41129, App
69	5	35.7	70	3	US-09-367-953B-47	Sequence 47, Appl
70	5	35.7	70	4	US-09-489-039A-12544	Sequence 12544, App
71	5	35.7	75	4	US-09-583-110-2743	Sequence 2743, App
72	5	35.7	77	3	US-09-134-001C-4812	Sequence 4812, App
73	5	35.7	78	4	US-09-248-796A-25922	Sequence 25922, App
74	5	35.7	79	4	US-09-513-999C-5308	Sequence 5308, App
75	5	35.7	81	4	US-09-134-000C-4348	Sequence 4348, App
76	5	35.7	81	4	US-09-621-976-4282	Sequence 4282, App
77	5	35.7	81	4	US-09-270-767-61813	Sequence 61813, App
78	5	35.7	82	4	US-09-270-767-61865	Sequence 61865, App
79	5	35.7	82	4	US-09-270-767-65881	Sequence 65881, App
80	5	35.7	87	4	US-09-178-093B-16	Sequence 16, Appl
81	5	35.7	87	4	US-09-248-796A-32366	Sequence 32366, App
82	5	35.7	91	4	US-09-621-976-5762	Sequence 5762, App
83	5	35.7	92	4	US-09-583-110-3504	Sequence 3504, App
84	5	35.7	100	4	US-09-732-210-451	Sequence 451, App
85	5	35.7	103	4	US-09-191-468-23	Sequence 23, Appl
86	5	35.7	103	4	US-09-191-468-25	Sequence 25, Appl
87	5	35.7	103	4	US-09-191-468-27	Sequence 27, Appl
88	5	35.7	103	4	US-09-191-468-29	Sequence 29, Appl
89	5	35.7	103	4	US-09-191-468-31	Sequence 31, Appl
90	5	35.7	103	4	US-09-191-468-33	Sequence 33, Appl
91	5	35.7	103	4	US-09-191-468-35	Sequence 35, Appl
92	5	35.7	106	4	US-09-191-468-37	Sequence 37, Appl
93	5	35.7	106	4	US-09-732-210-1463	Sequence 1463, App
94	5	35.7	108	1	US-08-436-463-17	Sequence 17, Appl
95	5	35.7	110	1	US-08-466-886-29	Sequence 29, Appl
96	5	35.7	110	3	US-08-688-871-62	Sequence 62, Appl
97	5	35.7	110	3	US-08-469-617-29	Sequence 29, Appl
98	5	35.7	110	4	US-08-469-617-29	Sequence 29, Appl
99	5	35.7	111	4	US-09-270-767-59945	Sequence 59945, App
100	5	35.7	116	4	US-09-270-767-33729	Sequence 33729, App

```

101 5 35.7 116 4 US-09-270-767-48946 Sequence 48946, A
102 5 35.7 121 4 US-09-759-112A-7 Sequence 7, Appl1
103 5 35.7 122 1 US-08-436-463-14 Sequence 14, Appl
104 5 35.7 123 4 US-09-248-796A-17814 Sequence 17814, A
105 5 35.7 124 4 US-09-543-681A-6095 Sequence 6095, Ap
106 5 35.7 125 4 US-09-690-454-155 Sequence 155, App
107 5 35.7 126 4 US-09-732-210-269 Sequence 269, App
108 5 35.7 127 4 US-09-732-210-729 Sequence 729, App
109 5 35.7 127 4 US-09-732-210-737 Sequence 737, App
110 5 35.7 128 4 US-09-583-110-4466 Sequence 4466, App
111 5 35.7 128 4 US-09-513-999C-7889 Sequence 7889, Ap
112 5 35.7 129 4 US-09-270-767-33971 Sequence 33971, A
113 5 35.7 129 4 US-09-270-767-49188 Sequence 49188, A
114 5 35.7 131 4 US-09-430-866A-4-4 Sequence 4, Appl1
115 5 35.7 131 4 US-09-248-796A-23124 Sequence 23124, A
116 5 35.7 132 4 US-09-732-210-589 Sequence 589, App
117 5 35.7 133 4 US-09-252-991A-24089 Sequence 24089, A
118 5 35.7 133 4 US-09-252-991A-29861 Sequence 29861, A
119 5 35.7 140 2 US-08-483-636-4 Sequence 4, Appl1
120 5 35.7 140 2 US-08-483-632-4 Sequence 4, Appl1
121 5 35.7 141 2 US-08-483-636-10 Sequence 10, Appl
122 5 35.7 141 2 US-08-483-632-10 Sequence 10, Appl
123 5 35.7 141 2 US-09-248-796A-19288 Sequence 19288, A
124 5 35.7 142 2 US-08-164-292B-20 Sequence 20, Appl
125 5 35.7 142 3 US-08-845-623-20 Sequence 20, Appl
126 5 35.7 142 3 US-08-815-927-20 Sequence 20, Appl
127 5 35.7 142 3 US-09-103-330-20 Sequence 20, Appl
128 5 35.7 142 3 US-09-435-242-20 Sequence 20, Appl
129 5 35.7 142 4 US-09-270-767-39503 Sequence 39503, A
130 5 35.7 142 4 US-09-270-767-40599 Sequence 40599, A
131 5 35.7 142 4 US-09-270-767-54720 Sequence 54720, A
132 5 35.7 142 4 US-09-270-767-55815 Sequence 55815, A
133 5 35.7 143 1 US-07-710-361-9 Sequence 9, Appl1
134 5 35.7 144 4 US-09-248-796A-14975 Sequence 14975, A
135 5 35.7 145 1 US-08-565-386-15 Sequence 15, Appl
136 5 35.7 149 4 US-09-270-767-39579 Sequence 39579, A
137 5 35.7 149 4 US-09-270-767-54796 Sequence 54796, A
138 5 35.7 150 4 US-09-252-991A-26074 Sequence 26074, A
139 5 35.7 150 4 US-09-621-976-4167 Sequence 4167, Ap
140 5 35.7 151 4 US-09-328-352-6429 Sequence 6429, Ap
141 5 35.7 154 3 US-09-387-418A-11 Sequence 11, Appl
142 5 35.7 155 4 US-09-270-767-31093 Sequence 31093, A
143 5 35.7 155 4 US-09-270-767-48310 Sequence 48310, A
144 5 35.7 156 4 US-09-621-976-4168 Sequence 4168, Ap
145 5 35.7 157 4 US-09-270-767-61769 Sequence 61769, A
146 5 35.7 165 4 US-09-489-039A-11468 Sequence 11468, A
147 5 35.7 167 4 US-09-270-767-44927 Sequence 44927, A
148 5 35.7 174 1 US-08-261-825-2 Sequence 2, Appl1
149 5 35.7 174 2 US-08-719-124-2 Sequence 2, Appl1
150 5 35.7 174 5 PCR-US95-07748A-2 Sequence 2, Appl1

```

## ALIGNMENTS

```

RESULT 1
US-09-134-000C-3462
Sequence 3462, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3462
LENGTH: 162
TYPE: PRT

```

```

ORGANISM: Enterococcus faecalis
US-09-134-000C-3462

```

```

Query Match 50.0%; Score 7; DB 4; Length 162;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 KITSVLS 12
DB 129 KITSVLS 135

```

```

RESULT 2
US-09-513-999C-4730
Sequence 4730, Application US/09513999C
Patent No. 6783961

```

```

GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J. B.
APPLICANT: Duclert, A. Y.

```

```

TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
Patent No. 6783961

```

```

FILE REFERENCE: 59.US2.REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24

```

```

PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26

```

```

NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 4730

```

```

LENGTH: 77
TYPE: PRT
ORGANISM: Homo sapiens
US-09-513-999C-4730

```

```

Query Match 42.9%; Score 6; DB 4; Length 77;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 4 KSKITS 9
DB 67 KSKITS 72

```

```

RESULT 3
US-09-107-532A-4535
Sequence 4535, Application US/09107532A
Patent No. 6583275

```

```

GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

```

```

NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA

```

```

ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>

```

```

SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998

```

```

APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997

```

```
ATTORNEY/AGENT INFORMATION:
NAME: Ariadello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4535:
SEQUENCE CHARACTERISTICS:
LENGTH: 133 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1..133
SEQUENCE DESCRIPTION: SEQ ID NO: 4535:
US-09-107-532A-4535

Query Match      42.9%; Score 6; DB 4; Length 133;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKITS 9
DB 120 KSKITS 125

RESULT 4
US-09-543-681A-4409
Sequence 4409, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 4409
LENGTH: 135
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-4409

Query Match      42.9%; Score 6; DB 4; Length 135;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 KITSVL 11
DB 3 KITSVL 8

RESULT 5
US-09-732-210-617
Sequence 617, Application US/09732210
Patent No. 6573361
GENERAL INFORMATION:
APPLICANT: Bunker, Greg J.
APPLICANT: Liang, Jihong
APPLICANT: Mitcanck, Cindy A.
APPLICANT: Seale, Jeffrey W.
APPLICANT: Wu, Yonnie S.
TITLE OF INVENTION: Anti-Fungal Proteins and Methods for Their Use
FILE REFERENCE: 38-21(15036)B
CURRENT APPLICATION NUMBER: US/09/732,210
```

```
CURRENT FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/169,513
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: US 60/169,340
PRIOR FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 1753
SEQ ID NO 617
LENGTH: 143
TYPE: PRT
ORGANISM: Methanococcus vannielii
US-09-732-210-617

Query Match      42.9%; Score 6; DB 4; Length 143;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSK 6
DB 1 MIRSK 6

RESULT 6
US-09-543-681A-7075
Sequence 7075, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 7075
LENGTH: 183
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-7075

Query Match      42.9%; Score 6; DB 4; Length 183;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSUUSF 13
DB 124 TSUUSF 129

RESULT 7
US-09-583-110-3590
Sequence 3590, Application US/09583110
Patent No. 6699703
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al.
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
FILE REFERENCE: PAT00-07A
CURRENT APPLICATION NUMBER: US/09/583,110
CURRENT FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/107,433
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/085,131
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: US 60/051,553
PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 5322
SEQ ID NO 3590
LENGTH: 238
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-583-110-3590
```

Query Match 42.9%; Score 6; DB 4; Length 238;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKITS 9  
|||||  
Db 162 KSKITS 167

RESULT 8  
US-09-107-532A-3734  
Sequence 3734, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Atinello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3734:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 270 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...270  
SEQUENCE DESCRIPTION: SEQ ID NO: 3734:  
US-09-107-532A-3734

Query Match 42.9%; Score 6; DB 4; Length 270;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
|||||  
Db 238 ITSVLS 243

RESULT 9  
US-09-248-796A-16584  
Sequence 16584, Application US/09248796A

Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 16584  
LENGTH: 275  
TYPE: PRT  
ORGANISM: Candida albicans  
US-09-248-796A-16584

Query Match 42.9%; Score 6; DB 4; Length 275;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
|||||  
Db 101 ITSVLS 106

RESULT 10  
US-09-489-039A-9786  
Sequence 9786, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 9786  
LENGTH: 311  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9786

Query Match 42.9%; Score 6; DB 4; Length 311;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
|||||  
Db 261 ITSVLS 266

RESULT 11  
US-09-526-597D-34  
Sequence 34, Application US/09526597D  
Patent No. 6710227  
GENERAL INFORMATION:  
APPLICANT: De Veylder, Lieven  
APPLICANT: De Almeida, Janice  
APPLICANT: Landrieu, Isabelle  
TITLE OF INVENTION: Cyclin-dependent kinase inhibitors and uses thereof  
FILE REFERENCE: 1187-2  
CURRENT APPLICATION NUMBER: US/09/526,597D  
CURRENT FILING DATE: 2000-03-16  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 34  
LENGTH: 327

TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-09-526-597D-34

Query Match 42.9% Score 6; DB 4; Length 327;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IRKSKI 7  
5 IRKSKI 10

RESULT 12  
US-09-107-532A-5546  
Sequence 5546, Application US/09107532A  
Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Walcham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Arinello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 5546:

SEQUENCE CHARACTERISTICS:

LENGTH: 344 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHEICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...344

SEQUENCE DESCRIPTION: SEQ ID NO: 5546:

US-09-107-532A-5546

Query Match 42.9% Score 6; DB 4; Length 344;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKITS 9  
163 KSKITS 168

RESULT 13  
US-08-700-359-20  
Sequence 20, Application US/08700359  
Patent No. 5766925

GENERAL INFORMATION:

APPLICANT: SUGIMOTO, MASAKAZU

APPLICANT: USUDA, YOSHIHIRO

APPLICANT: SUZUKI, TOMOKO

APPLICANT: TANAKA, AKIKO

APPLICANT: MATSUI, HIROSHI

TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: MS-DOS TEXT EDITOR

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/700,359

FILING DATE: 08-OCT-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP-35019

FILING DATE: 04-MAR-1994

ATTORNEY/AGENT INFORMATION:

NAME: NORMAN F. OBLON

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 10-819-0 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 433 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-700-359-20

Query Match 42.9% Score 6; DB 1; Length 433;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
363 SKITSV 368

RESULT 14

US-09-019-095A-24

Sequence 24, Application US/09019095A

Patent No. 6287858

GENERAL INFORMATION:

APPLICANT: D'Andrea, Alan D.

TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate

FILE REFERENCE: DFC1-435P2A2

CURRENT APPLICATION NUMBER: US/09/019,095A

CURRENT FILING DATE: 1998-02-05

PRIOR APPLICATION NUMBER: PCT/US96/12884

PRIOR FILING DATE: 1996-08-07

PRIOR APPLICATION NUMBER: US 60/002,066

PRIOR FILING DATE: 1995-08-09

PRIOR APPLICATION NUMBER: US 60/019,787

PRIOR FILING DATE: 1996-06-14

NUMBER OF SEQ ID NOS: 51



```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 494
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-019-095A-24
```

```
Query Match          42.9%; Score 6; DB 3; Length 494;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      7 ITSVLS 12
        |||||
Db      324 ITSVLS 329
```

```
RESULT 15
US-08-945-771-2
; Sequence 2, Application US/08945771
; Patent No. 6465623
; GENERAL INFORMATION:
; APPLICANT: Daly, Roger J
; APPLICANT: Sutherland, Robert L
; TITLE OF INVENTION: GDU, A novel signalling protein
; FILE REFERENCE: 273402001700
; CURRENT APPLICATION NUMBER: US/08/945, 771
; CURRENT FILING DATE: 1998-04-22
; EARLIER APPLICATION NUMBER: PCT/US96/00258
; EARLIER FILING DATE: 1996-MAY-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-945-771-2
```

```
Query Match          42.9%; Score 6; DB 4; Length 540;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      7 ITSVLS 12
        |||||
Db      90 ITSVLS 95
```

```
RESULT 16
US-09-371-338-7
; Sequence 7, Application US/09371338
; Patent No. 6613959
; GENERAL INFORMATION:
; APPLICANT: Sheen, Jen
; APPLICANT: Kourtun, Yelena V.
; APPLICANT: Chiu, Wan-Ling
; TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING A MAPKKK
; FILE REFERENCE: 00786/366002
; CURRENT APPLICATION NUMBER: US/09/371,338
; CURRENT FILING DATE: 1999-08-10
; PRIOR APPLICATION NUMBER: 60/095,938
; PRIOR FILING DATE: 1998-08-10
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 661
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-371-338-7
```

```
Query Match          42.9%; Score 6; DB 4; Length 661;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 IRKSKI 7
        |||||
Db      36 IRKSKI 41
```

```
RESULT 17
US-09-134-001C-3586
; Sequence 3586, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3586
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3586
```

```
Query Match          42.9%; Score 6; DB 3; Length 739;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      8 TSVLSF 13
        |||||
Db      446 TSVLSF 451
```

```
RESULT 18
US-09-543-681A-6407
; Sequence 6407, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6407
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6407
```

```
Query Match          42.9%; Score 6; DB 4; Length 967;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      7 ITSVLS 12
        |||||
Db      781 ITSVLS 786
```

```
RESULT 19
US-09-177-249-124
; Sequence 124, Application US/09177249
; Patent No. 6223064
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Ohad, Nir
; APPLICANT: Kiyosue, Tomohiro
```

APPLICANT: Yadegari, Ramlin  
APPLICANT: Margosian, Linda  
APPLICANT: Harada, John  
APPLICANT: Goldberg, Robert B.  
TITLE OF INVENTION: Regents of the University of California  
TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit  
FILE REFERENCE: 023070-086120US  
CURRENT APPLICATION NUMBER: US/09/177,249  
CURRENT FILING DATE: 1998-10-22  
EARLIER APPLICATION NUMBER: US 09/071,838  
EARLIER FILING DATE: 1998-05-01  
NUMBER OF SEQ ID NOS: 324  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 124  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Arabidopsis sp.  
US-09-177-249-124

Query Match 35.7%; Score 5; DB 3; length 12;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVL 11  
Db 5 ITSVL 9

RESULT 20  
US-09-522-666-30  
Sequence 30, Application US/09522666  
Patent No. 6331167  
GENERAL INFORMATION:  
APPLICANT: Shuey, David  
APPLICANT: Quinet, Elaine  
TITLE OF INVENTION: Methods and Reagents for Identifying Inhibitors of  
TITLE OF INVENTION: Protocols of Membrane-Associated Proteins  
FILE REFERENCE: 6-00  
CURRENT APPLICATION NUMBER: US/09/522,666  
CURRENT FILING DATE: 2000-03-10  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 30  
LENGTH: 13  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: amino acid  
OTHER INFORMATION: sequence encompassing Site-1 protease cleavage  
OTHER INFORMATION: site within SREPB-2  
US-09-522-666-30

Query Match 35.7%; Score 5; DB 3; length 13;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSF 13  
Db 8 SVLSF 12

RESULT 21  
US-08-268-251-5  
Sequence 5, Application US/08268251  
Patent No. 5585475  
GENERAL INFORMATION:  
APPLICANT: Jamieson, Gordon A  
APPLICANT: Dedman, John R  
APPLICANT: Kaetzel, Marcia A  
TITLE OF INVENTION: Calmodulin-Binding Peptides  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/268,251  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/831,219  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D  
REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 272,001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2706  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-268-251-5

Query Match 35.7%; Score 5; DB 1; length 15;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSVLS 12  
Db 9 TSVLS 13

RESULT 22  
PCT-US93-01112-5  
Sequence 5, Application PC/TUS9301112  
GENERAL INFORMATION:  
APPLICANT: Jamieson, Gordon A  
APPLICANT: Dedman, John R  
APPLICANT: Kaetzel, Marcia A  
TITLE OF INVENTION: Calmodulin-Binding Peptides  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01112  
FILING DATE: 19930208  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/831,219  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D

REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 272,001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2706  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US93-01112-5

Query Match 35.7%; Score 5; DB 5; Length 15;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSLS 12  
Db 9 TSLS 13

RESULT 23  
US-09-360-237-55  
Sequence 55, Application US/09360237  
Patent No. 6322962  
GENERAL INFORMATION:  
APPLICANT: BROWN, MICHAEL S.  
APPLICANT: CHENG, DONG  
APPLICANT: ESPENSHADE, PETER J.  
APPLICANT: GOLDSTEIN, JOSEPH L.  
APPLICANT: RAMSON, ROBERT B.  
APPLICANT: SAKAI, JIRO  
TITLE OF INVENTION: STEROL-REGULATED SITE-1 PROTEASE AND ASSAYS OF  
TITLE OF INVENTION: MODULATORS THEREOF  
FILE REFERENCE: UTXD:567  
CURRENT APPLICATION NUMBER: US/09/360,237  
CURRENT FILING DATE: 1999-07-23  
EARLIER APPLICATION NUMBER: 60/096,571  
EARLIER FILING DATE: 1998-08-14  
NUMBER OF SEQ ID NOS: 60  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 55  
LENGTH: 16  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-360-237-55

Query Match 35.7%; Score 5; DB 3; Length 16;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSF 13  
Db 7 SVLSF 11

RESULT 24  
US-09-556-877-251  
Sequence 251, Application US/09556877  
Patent No. 6432916  
GENERAL INFORMATION:  
APPLICANT: Probst, Peter  
APPLICANT: Bhatia, Ajay  
APPLICANT: Skeiky, Yasir  
APPLICANT: Fling, Steve  
APPLICANT: Maisonneuve, Jeff  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

FILE REFERENCE: 210121.469C5  
CURRENT APPLICATION NUMBER: US/09/556,877  
CURRENT FILING DATE: 2000-04-19  
NUMBER OF SEQ ID NOS: 305  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 251  
LENGTH: 16  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Made in a lab  
US-09-556-877-251

Query Match 35.7%; Score 5; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKIT 8  
Db 3 KSKIT 7

RESULT 25  
US-09-620-412C-251  
Sequence 251, Application US/09620412C  
Patent No. 6448234  
GENERAL INFORMATION:  
APPLICANT: Steven P. Fling  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
FILE REFERENCE: 210121.469C7  
CURRENT APPLICATION NUMBER: US/09/620,412C  
CURRENT FILING DATE: 2000-07-20  
NUMBER OF SEQ ID NOS: 363  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 251  
LENGTH: 16  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Made in a lab  
US-09-620-412C-251

Query Match 35.7%; Score 5; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKIT 8  
Db 3 KSKIT 7

RESULT 26  
US-09-598-419-251  
Sequence 251, Application US/09598419  
Patent No. 6565856  
GENERAL INFORMATION:  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Scholler, John  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
FILE REFERENCE: 210121.469C6  
CURRENT APPLICATION NUMBER: US/09/598,419  
CURRENT FILING DATE: 2000-06-20  
NUMBER OF SEQ ID NOS: 357  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 251  
LENGTH: 16  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Made in a lab  
US-09-598-419-251

Query Match 35.7%; Score 5; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKIT 8  
|||  
Db 3 KSKIT 7

RESULT 27  
US-09-680-571A-83  
; Sequence 83, Application US/09680571A  
; Patent No. 6649593  
; GENERAL INFORMATION:

; APPLICANT: Jaen, Juan C.  
; APPLICANT: Li, Leping  
; APPLICANT: Brown, Michael S.  
; APPLICANT: Goldstein, Joseph L.  
; APPLICANT: Cheng, Dong  
; APPLICANT: Tularik Inc.  
; APPLICANT: Board of Regents, The University of Texas System  
; APPLICANT: Southwestern Medical Center  
; TITLE OF INVENTION: Modulators of SREBP Processing  
; FILE REFERENCE: 018781-002810US  
; CURRENT APPLICATION NUMBER: US/09/680,571A  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/159,236  
; PRIOR FILING DATE: 1999-10-13  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 83  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: typical Site-1  
; OTHER INFORMATION: protease (S1P) peptide cleavage substrate  
US-09-680-571A-83

Query Match 35.7%; Score 5; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSF 13  
|||  
Db 7 SVLSF 11

RESULT 28  
US-09-680-571A-97  
; Sequence 97, Application US/09680571A  
; Patent No. 6649593  
; GENERAL INFORMATION:

; APPLICANT: Jaen, Juan C.  
; APPLICANT: Li, Leping  
; APPLICANT: Brown, Michael S.  
; APPLICANT: Goldstein, Joseph L.  
; APPLICANT: Cheng, Dong  
; APPLICANT: Tularik Inc.  
; APPLICANT: Board of Regents, The University of Texas System  
; APPLICANT: Southwestern Medical Center  
; TITLE OF INVENTION: Modulators of SREBP Processing  
; FILE REFERENCE: 018781-002810US  
; CURRENT APPLICATION NUMBER: US/09/680,571A  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/159,236  
; PRIOR FILING DATE: 1999-10-13  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 97  
; LENGTH: 16  
; TYPE: PRT

; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthesized  
; OTHER INFORMATION: peptide  
US-09-680-571A-97

Query Match 35.7%; Score 5; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSF 13  
|||  
Db 7 SVLSF 11

RESULT 29  
US-09-556-877-248  
; Sequence 248, Application US/09556877  
; Patent No. 6432916  
; GENERAL INFORMATION:

; APPLICANT: Probst, Peter  
; APPLICANT: Bhacia, Ajay  
; APPLICANT: Skeiky, Yasir  
; APPLICANT: Fling, Steve  
; APPLICANT: Maisonneuve, Jeff  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; FILE REFERENCE: 210121.469C5  
; CURRENT APPLICATION NUMBER: US/09/556,877  
; PRIOR FILING DATE: 2000-04-19  
; NUMBER OF SEQ ID NOS: 305  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 248  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Made in a lab  
US-09-556-877-248

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKIT 8  
|||  
Db 12 KSKIT 16

RESULT 30  
US-09-556-877-249  
; Sequence 249, Application US/09556877  
; Patent No. 6432916  
; GENERAL INFORMATION:

; APPLICANT: Probst, Peter  
; APPLICANT: Bhacia, Ajay  
; APPLICANT: Skeiky, Yasir  
; APPLICANT: Fling, Steve  
; APPLICANT: Maisonneuve, Jeff  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; FILE REFERENCE: 210121.469C5  
; CURRENT APPLICATION NUMBER: US/09/556,877  
; PRIOR FILING DATE: 2000-04-19  
; NUMBER OF SEQ ID NOS: 305  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 249  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Made in a lab  
US-09-556-877-249

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
      |||||  
Db 7 KSKIT 11

RESULT 31  
US-09-556-877-250  
; Sequence 250, Application US/09556877  
; Patent No. 6432916  
; GENERAL INFORMATION:  
; APPLICANT: Probst, Peter  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Skeiky, Yasir  
; APPLICANT: Fling, Steve  
; APPLICANT: Maisonneuve, Jeff  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C5  
; CURRENT APPLICATION NUMBER: US/09/556,877  
; CURRENT FILING DATE: 2000-04-19  
; NUMBER OF SEQ ID NOS: 305  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 250  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Made in a lab  
US-09-556-877-250

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
      |||||  
Db 3 KSKIT 7

RESULT 32  
US-09-620-412C-248  
; Sequence 248, Application US/09620412C  
; Patent No. 6448234  
; GENERAL INFORMATION:  
; APPLICANT: Steven P. Fling  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C7  
; CURRENT APPLICATION NUMBER: US/09/620,412C  
; CURRENT FILING DATE: 2000-07-20  
; NUMBER OF SEQ ID NOS: 363  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 248  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Made in a lab  
US-09-620-412C-248

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
      |||||  
Db 12 KSKIT 16

RESULT 33  
US-09-620-412C-249  
; Sequence 249, Application US/09620412C  
; Patent No. 6448234  
; GENERAL INFORMATION:  
; APPLICANT: Steven P. Fling  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C7  
; CURRENT APPLICATION NUMBER: US/09/620,412C  
; CURRENT FILING DATE: 2000-07-20  
; NUMBER OF SEQ ID NOS: 363  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 249  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Made in a lab  
US-09-620-412C-249

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
      |||||  
Db 7 KSKIT 11

RESULT 34  
US-09-620-412C-250  
; Sequence 250, Application US/09620412C  
; Patent No. 6448234  
; GENERAL INFORMATION:  
; APPLICANT: Steven P. Fling  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C7  
; CURRENT APPLICATION NUMBER: US/09/620,412C  
; CURRENT FILING DATE: 2000-07-20  
; NUMBER OF SEQ ID NOS: 363  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 250  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Made in a lab  
US-09-620-412C-250

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
      |||||  
Db 3 KSKIT 7

RESULT 35  
US-09-598-419-248  
; Sequence 248, Application US/09598419  
; Patent No. 6565856  
; GENERAL INFORMATION:  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Scholler, John  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C6  
; CURRENT APPLICATION NUMBER: US/09/598,419  
; CURRENT FILING DATE: 2000-06-20

NUMBER OF SEQ ID NOS: 357  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 248  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Made in a lab  
US-09-598-419-248

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
Db 12 KSKIT 16

RESULT 36  
US-09-598-419-249  
Sequence 249, Application US/09598419  
Patent No. 6565856  
GENERAL INFORMATION:  
APPLICANT: Skelky, Yasir A.W.  
APPLICANT: Scholler, John  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
FILE REFERENCE: 210121.469C6  
CURRENT APPLICATION NUMBER: US/09/598,419  
CURRENT FILING DATE: 2000-06-20  
NUMBER OF SEQ ID NOS: 357  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 249  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Made in a lab  
US-09-598-419-249

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
Db 7 KSKIT 11

RESULT 37  
US-09-598-419-250  
Sequence 250, Application US/09598419  
Patent No. 6565856  
GENERAL INFORMATION:  
APPLICANT: Skelky, Yasir A.W.  
APPLICANT: Scholler, John  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
FILE REFERENCE: 210121.469C6  
CURRENT APPLICATION NUMBER: US/09/598,419  
CURRENT FILING DATE: 2000-06-20  
NUMBER OF SEQ ID NOS: 357  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 250  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Made in a lab  
US-09-598-419-250

Query Match 35.7%; Score 5; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
Db 3 KSKIT 7

RESULT 38  
US-09-834-759-545  
Sequence 545, Application US/09834759  
Patent No. 6680197  
GENERAL INFORMATION:  
APPLICANT: Jiang, Yugui  
APPLICANT: Dillon, Devin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Xu, Jiangchun  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Hepler, William T.  
APPLICANT: Henderson, Robert A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
FILE REFERENCE: 210121.470C9  
CURRENT APPLICATION NUMBER: US/09/834,759  
CURRENT FILING DATE: 2001-04-13  
NUMBER OF SEQ ID NOS: 547  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 545  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-834-759-545

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
Db 3 KSKIT 7

RESULT 39  
US-08-268-251-35  
Sequence 35, Application US/08268251  
Patent No. 3585475  
GENERAL INFORMATION:  
APPLICANT: Jamieson, Gordon A  
APPLICANT: Dedman, John R  
APPLICANT: Kaetzel, Marcia A  
TITLE OF INVENTION: Calmodulin-Binding Peptides  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/268,251  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/831,219  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D

REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 272.001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2706  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-268-251-35

Query Match 35.7%; Score 5; DB 1; Length 23;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 TSVLS 12  
|||  
Db 11 TSVLS 15

RESULT 40  
PCT-US93-01112-35  
Sequence 35, Application PC/TUS9301112  
GENERAL INFORMATION:  
APPLICANT: Jamieson, Gordon A  
APPLICANT: Dedman, John R  
APPLICANT: Kaetzl, Marcia A  
TITLE OF INVENTION: Calmodulin-Binding Peptides  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01112  
FILING DATE: 19930208  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/831,219  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D  
REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 272.001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2706  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US93-01112-35

Query Match 35.7%; Score 5; DB 5; Length 23;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 TSVLS 12  
|||  
Db 11 TSVLS 15

Db 11 TSVLS 15

Search completed: December 30, 2004, 15:37:58  
Job time: 42 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2004, 15:34:09 ; Search time 144 Seconds  
(without alignments)  
34.974 Million cell updates/sec

Title: US-10-719-385-20  
Perfect score: 14  
Sequence: 1 MIRSKITSVLSFC 14

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1599051 seqs, 359727711 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1599051

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 150 summaries

Database :

Published Applications MA:\*

- 1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubppaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubppaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubppaa/US09C\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubppaa/US10D\_PUBCOMB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubppaa/US11\_NEW\_PUB.pep:\*
- 19: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*
- 20: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	14	100.0	14 17 US-10-719-385-20	Sequence 20, Appl1
2	14	100.0	1752 17 US-10-719-385-5	Sequence 5, Appl1
3	14	100.0	1753 17 US-10-719-385-2	Sequence 2, Appl1
4	14	100.0	1753 17 US-10-719-385-3	Sequence 3, Appl1
5	14	100.0	1753 17 US-10-719-385-4	Sequence 4, Appl1
6	14	100.0	1753 17 US-10-719-385-6	Sequence 6, Appl1
7	14	100.0	1753 17 US-10-719-385-7	Sequence 7, Appl1
8	14	100.0	1753 17 US-10-719-385-8	Sequence 8, Appl1
9	14	100.0	1753 17 US-10-719-385-9	Sequence 9, Appl1
10	14	100.0	1753 17 US-10-719-385-10	Sequence 10, Appl1
11	14	100.0	1753 17 US-10-719-385-11	Sequence 11, Appl1
12	14	100.0	1753 17 US-10-719-385-12	Sequence 12, Appl1
13	14	100.0	1753 17 US-10-719-385-13	Sequence 13, Appl1

14	14	100.0	1753 17 US-10-719-385-14	Sequence 14, Appl1
15	14	100.0	1753 17 US-10-719-385-15	Sequence 15, Appl1
16	14	100.0	1753 17 US-10-719-385-16	Sequence 16, Appl1
17	14	100.0	1753 17 US-10-719-385-17	Sequence 17, Appl1
18	14	100.0	1753 17 US-10-719-385-18	Sequence 18, Appl1
19	14	100.0	1753 17 US-10-719-385-19	Sequence 19, Appl1
20	14	100.0	1753 17 US-10-719-385-20	Sequence 20, Appl1
21	7	50.0	285 16 US-10-424-599-148727	Sequence 148727, A
22	7	50.0	290 15 US-10-781-014-694	Sequence 694, App
23	7	50.0	363 9 US-10-424-599-158692	Sequence 158692, A
24	7	50.0	431 15 US-10-282-122A-57038	Sequence 57038, A
25	7	50.0	1019 9 US-09-738-626-6499	Sequence 6499, Ap
26	7	50.0	1019 9 US-10-380-055-12	Sequence 12, Appl1
27	7	50.0	1065 15 US-10-424-599-148728	Sequence 148728, A
28	7	50.0	12 14 US-10-775-869-10	Sequence 10, Appl1
29	6	42.9	12 14 US-10-366-493-10	Sequence 10, Appl1
30	6	42.9	17 14 US-10-029-347-10	Sequence 10, Appl1
31	6	42.9	148 16 US-10-437-963-111764	Sequence 111764, A
32	6	42.9	53 15 US-10-424-599-1267615	Sequence 11764, A
33	6	42.9	56 17 US-10-425-115-185349	Sequence 185349, A
34	6	42.9	60 16 US-10-437-963-107926	Sequence 185349, A
35	6	42.9	69 15 US-10-424-599-284793	Sequence 284793, A
36	6	42.9	82 15 US-10-424-599-143794	Sequence 143794, A
37	6	42.9	100 15 US-10-424-599-217068	Sequence 217068, A
38	6	42.9	100 15 US-10-425-115-307173	Sequence 307173, A
39	6	42.9	111 15 US-10-425-115-54103	Sequence 54103, A
40	6	42.9	130 15 US-10-282-122A-45474	Sequence 45474, A
41	6	42.9	151 15 US-10-425-114-44778	Sequence 44778, A
42	6	42.9	178 16 US-10-767-701-38935	Sequence 38935, A
43	6	42.9	184 17 US-10-370-838-9	Sequence 9, Appl1
44	6	42.9	184 17 US-10-789-433-12	Sequence 12, Appl1
45	6	42.9	186 15 US-10-270-838-8	Sequence 8, Appl1
46	6	42.9	186 15 US-10-789-433-14	Sequence 14, Appl1
47	6	42.9	187 17 US-10-789-433-9	Sequence 9, Appl1
48	6	42.9	187 17 US-10-789-433-10	Sequence 10, Appl1
49	6	42.9	187 17 US-10-789-433-11	Sequence 11, Appl1
50	6	42.9	198 16 US-10-437-963-129082	Sequence 129082, A
51	6	42.9	258 17 US-10-739-930-9122	Sequence 9122, Ap
52	6	42.9	289 15 US-10-451-139-17	Sequence 17, Appl1
53	6	42.9	302 17 US-10-149-826-11	Sequence 11, Appl1
54	6	42.9	304 17 US-10-425-115-270799	Sequence 270799, A
55	6	42.9	311 9 US-09-886-055-15	Sequence 15, Appl1
56	6	42.9	311 15 US-09-804-291-15	Sequence 15, Appl1
57	6	42.9	311 15 US-10-343-650A-402	Sequence 402, App
58	6	42.9	314 14 US-10-025-806-18	Sequence 18, Appl1
59	6	42.9	317 15 US-09-965-821-60	Sequence 60, Appl1
60	6	42.9	317 15 US-10-407-866-60	Sequence 60, Appl1
61	6	42.9	317 16 US-10-781-294-60	Sequence 60, Appl1
62	6	42.9	321 15 US-10-407-866-107	Sequence 107, App
63	6	42.9	321 15 US-10-264-237-2549	Sequence 2549, Ap
64	6	42.9	326 14 US-10-160-764-5	Sequence 5, Appl1
65	6	42.9	326 15 US-10-229-541A-11	Sequence 11, Appl1
66	6	42.9	326 17 US-10-773-529-20	Sequence 20, Appl1
67	6	42.9	327 15 US-10-688-291-34	Sequence 34, Appl1
68	6	42.9	360 10 US-09-372-348-9	Sequence 9, Appl1
69	6	42.9	374 15 US-09-372-321-4	Sequence 4, Appl1
70	6	42.9	384 17 US-10-128-558-205	Sequence 205, App
71	6	42.9	387 16 US-10-408-65A-2232	Sequence 2232, Ap
72	6	42.9	403 16 US-10-469-864-2	Sequence 2, Appl1
73	6	42.9	404 16 US-10-474-776-724	Sequence 724, App
74	6	42.9	431 15 US-10-282-122A-57773	Sequence 57773, A
75	6	42.9	433 14 US-10-369-493-17137	Sequence 17137, A
76	6	42.9	433 14 US-10-369-493-23280	Sequence 23280, A
77	6	42.9	458 14 US-10-369-493-3910	Sequence 3910, Ap
78	6	42.9	497 14 US-10-371-905A-28	Sequence 28, Appl1
79	6	42.9	497 14 US-10-372-348-11	Sequence 11, Appl1
80	6	42.9	530 14 US-10-107-695B-2	Sequence 2, Appl1
81	6	42.9	530 14 US-10-371-905A-6	Sequence 6, Appl1
82	6	42.9	530 14 US-10-371-905A-12	Sequence 12, Appl1
83	6	42.9	530 14 US-10-371-905A-16	Sequence 16, Appl1
84	6	42.9	530 14 US-10-371-905A-18	Sequence 18, Appl1
85	6	42.9	530 14 US-10-371-905A-30	Sequence 30, Appl1
86	6	42.9	530 14 US-10-371-905A-32	Sequence 32, Appl1

87	6	42.9	530	14	US-10-371-905A-34	Sequence 34, Appl
88	6	42.9	530	14	US-10-371-905A-36	Sequence 36, Appl
89	6	42.9	540	14	US-08-945-771-2	Sequence 2, Appl1
90	6	42.9	540	14	US-10-242-332-2	Sequence 2, Appl1
91	6	42.9	540	14	US-10-323-001-2	Sequence 2, Appl1
92	6	42.9	558	15	US-10-282-122A-53651	Sequence 53651, A
93	6	42.9	565	14	US-10-371-905A-20	Sequence 20, Appl
94	6	42.9	574	14	US-10-371-905A-14	Sequence 14, Appl
95	6	42.9	578	15	US-10-428-275-14	Sequence 14, Appl
96	6	42.9	584	14	US-10-029-386-32446	Sequence 32446, A
97	6	42.9	597	15	US-10-112-944-770	Sequence 770, App
98	6	42.9	597	15	US-10-276-774-2105	Sequence 2105, Ap
99	6	42.9	604	17	US-10-425-115-270800	Sequence 270800, A
100	6	42.9	618	15	US-10-425-114-62636	Sequence 62636, A
101	6	42.9	625	14	US-10-029-347-2	Sequence 2, Appl1
102	6	42.9	633	14	US-10-371-905A-24	Sequence 24, Appl
103	6	42.9	637	15	US-10-282-122A-48281	Sequence 48281, A
104	6	42.9	642	17	US-10-425-115-270798	Sequence 270798, A
105	6	42.9	650	15	US-10-282-122A-66928	Sequence 66928, A
106	6	42.9	651	9	US-09-815-242-11022	Sequence 11022, A
107	6	42.9	651	9	US-10-282-122A-58169	Sequence 58169, A
108	6	42.9	659	15	US-10-407-866-91	Sequence 91, Appl
109	6	42.9	661	15	US-10-643-434-7	Sequence 7, Appl1
110	6	42.9	676	15	US-10-092-900A-70	Sequence 70, Appl
111	6	42.9	693	15	US-10-282-122A-78018	Sequence 78018, A
112	6	42.9	750	16	US-10-437-963-129080	Sequence 129080, A
113	6	42.9	757	16	US-10-437-963-146837	Sequence 146837, A
114	6	42.9	767	15	US-10-424-599-198727	Sequence 198727, A
115	6	42.9	779	16	US-10-437-963-133669	Sequence 133669, A
116	6	42.9	782	15	US-10-282-122A-43733	Sequence 43733, A
117	6	42.9	804	16	US-10-437-963-203648	Sequence 203648, A
118	6	42.9	831	17	US-10-723-860-2937	Sequence 2937, Ap
119	6	42.9	862	10	US-09-825-751A-65	Sequence 65, Appl
120	6	42.9	891	15	US-10-112-944-312	Sequence 312, App
121	6	42.9	891	17	US-10-149-826-2	Sequence 2, Appl1
122	6	42.9	892	14	US-10-446-046-2	Sequence 2, Appl1
123	6	42.9	902	15	US-10-407-866-66	Sequence 66, Appl
124	6	42.9	909	14	US-10-369-493-2310	Sequence 2310, A
125	6	42.9	947	10	US-09-865-621-18	Sequence 18, Appl
126	6	42.9	947	15	US-10-407-866-18	Sequence 18, Appl
127	6	42.9	947	16	US-10-781-294-18	Sequence 18, Appl
128	6	42.9	1000	14	US-10-437-963-187544	Sequence 187544, A
129	6	42.9	1016	14	US-10-371-905A-4	Sequence 4, Appl1
130	6	42.9	1021	14	US-10-371-905A-8	Sequence 8, Appl1
131	5	35.7	9	10	US-09-935-384-211	Sequence 211, App
132	5	35.7	9	10	US-09-935-384-322	Sequence 322, App
133	5	35.7	9	10	US-09-935-384-326	Sequence 326, App
134	5	35.7	9	10	US-09-935-384-416	Sequence 416, App
135	5	35.7	9	10	US-09-935-384-527	Sequence 527, App
136	5	35.7	9	10	US-09-935-384-610	Sequence 610, App
137	5	35.7	10	10	US-08-572-404B-3732	Sequence 3732, Ap
138	5	35.7	10	10	US-08-572-404B-3733	Sequence 3733, Ap
139	5	35.7	10	10	US-09-572-404B-3734	Sequence 3734, Ap
140	5	35.7	10	10	US-09-935-384-161	Sequence 161, App
141	5	35.7	10	10	US-09-935-384-175	Sequence 175, App
142	5	35.7	10	10	US-09-935-384-372	Sequence 372, App
143	5	35.7	10	10	US-09-935-384-396	Sequence 396, App
144	5	35.7	10	10	US-09-935-384-593	Sequence 593, App
145	5	35.7	10	10	US-09-935-384-692	Sequence 692, App
146	5	35.7	12	9	US-09-071-838-124	Sequence 124, App
147	5	35.7	12	9	US-10-213-512-124	Sequence 124, App
148	5	35.7	14	17	US-10-413-600-28	Sequence 28, Appl
149	5	35.7	16	9	US-09-841-133-251	Sequence 251, App
150	5	35.7	16	17	US-10-872-155-251	Sequence 251, App

## ALIGNMENTS

GENERAL INFORMATION:  
 APPLICANT: O'Toole et al.  
 TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 FILE REFERENCE: 22058-582  
 CURRENT APPLICATION NUMBER: US/10/719,385  
 CURRENT FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: PCT/US03/37339  
 PRIOR FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: 60/428,094  
 PRIOR FILING DATE: 2002-11-21  
 NUMBER OF SEQ ID NOS: 26  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 20  
 LENGTH: 14  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-719-385-20

Query Match  
 Best Local Similarity 100.0%; Score 14; DB 17; Length 14;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSPC 14  
 DB 1 MIRSKITSVLSPC 14

RESULT 2  
 US-10-719-385-5  
 Sequence 5, Application US/10719385  
 Publication No. US20040209284A1  
 GENERAL INFORMATION:

APPLICANT: O'Toole et al.  
 TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 FILE REFERENCE: 22058-582  
 CURRENT APPLICATION NUMBER: US/10/719,385  
 CURRENT FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: PCT/US03/37339  
 PRIOR FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: 60/428,094  
 PRIOR FILING DATE: 2002-11-21  
 NUMBER OF SEQ ID NOS: 26  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 5  
 LENGTH: 1752  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-719-385-5

Query Match  
 Best Local Similarity 100.0%; Score 14; DB 17; Length 1752;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSPC 14  
 DB 1 MIRSKITSVLSPC 14

RESULT 3  
 US-10-719-385-2  
 Sequence 2, Application US/10719385  
 Publication No. US20040209284A1  
 GENERAL INFORMATION:

APPLICANT: O'Toole et al.  
 TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 FILE REFERENCE: 22058-582  
 CURRENT APPLICATION NUMBER: US/10/719,385  
 CURRENT FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: PCT/US03/37339  
 PRIOR FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: 60/428,094  
 PRIOR FILING DATE: 2002-11-21  
 NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 2  
 LENGTH: 1753  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-719-385-2

Query Match 100.0%; Score 14; DB 17; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

RESULT 4  
 US-10-719-385-3

Sequence 3, Application US/10719385  
 Publication No. US20040209284A1  
 GENERAL INFORMATION:

APPLICANT: O'Toole et al.  
 TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 FILE REFERENCE: 22058-582  
 CURRENT APPLICATION NUMBER: US/10/719,385  
 CURRENT FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: PCT/US03/37339  
 PRIOR FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: 60/428,094  
 PRIOR FILING DATE: 2002-11-21  
 NUMBER OF SEQ ID NOS: 26  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 3  
 LENGTH: 1753  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-719-385-3

Query Match 100.0%; Score 14; DB 17; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

RESULT 5  
 US-10-719-385-4

Sequence 4, Application US/10719385  
 Publication No. US20040209284A1  
 GENERAL INFORMATION:

APPLICANT: O'Toole et al.  
 TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 FILE REFERENCE: 22058-582  
 CURRENT APPLICATION NUMBER: US/10/719,385  
 CURRENT FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: PCT/US03/37339  
 PRIOR FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: 60/428,094  
 PRIOR FILING DATE: 2002-11-21  
 NUMBER OF SEQ ID NOS: 26  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 4  
 LENGTH: 1753  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-719-385-4

Query Match 100.0%; Score 14; DB 17; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

RESULT 6  
 US-10-719-385-6

Sequence 6, Application US/10719385  
 Publication No. US20040209284A1  
 GENERAL INFORMATION:  
 APPLICANT: O'Toole et al.  
 TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 FILE REFERENCE: 22058-582  
 CURRENT APPLICATION NUMBER: US/10/719,385  
 CURRENT FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: PCT/US03/37339  
 PRIOR FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: 60/428,094  
 PRIOR FILING DATE: 2002-11-21  
 NUMBER OF SEQ ID NOS: 26  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 6  
 LENGTH: 1753  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-719-385-6

Query Match 100.0%; Score 14; DB 17; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

RESULT 7  
 US-10-719-385-7

Sequence 7, Application US/10719385  
 Publication No. US20040209284A1  
 GENERAL INFORMATION:  
 APPLICANT: O'Toole et al.  
 TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 FILE REFERENCE: 22058-582  
 CURRENT APPLICATION NUMBER: US/10/719,385  
 CURRENT FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: PCT/US03/37339  
 PRIOR FILING DATE: 2003-11-21  
 PRIOR APPLICATION NUMBER: 60/428,094  
 PRIOR FILING DATE: 2002-11-21  
 NUMBER OF SEQ ID NOS: 26  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 7  
 LENGTH: 1753  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-719-385-7

Query Match 100.0%; Score 14; DB 17; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

RESULT 8  
 US-10-719-385-8

Sequence 8, Application US/10719385  
 Publication No. US20040209284A1  
 GENERAL INFORMATION:  
 APPLICANT: O'Toole et al.

TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
FILE REFERENCE: 22058-582  
CURRENT APPLICATION NUMBER: US/10/719,385  
CURRENT FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: PCT/US03/37339  
PRIOR FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: 60/428,094  
PRIOR FILING DATE: 2002-11-21  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 1753  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-719-385-8

Query Match 100.0%; Score 14; DB 17; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
DB 1 MIRSKITSVLSFC 14

RESULT 9  
US-10-719-385-9

Sequence 9, Application US/10719385  
Publication No. US20040209284A1  
GENERAL INFORMATION:

APPLICANT: O'Toole et al.

TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis

FILE REFERENCE: 22058-582

CURRENT APPLICATION NUMBER: US/10/719,385

CURRENT FILING DATE: 2003-11-21

PRIOR APPLICATION NUMBER: PCT/US03/37339

PRIOR FILING DATE: 2003-11-21

PRIOR APPLICATION NUMBER: 60/428,094

PRIOR FILING DATE: 2002-11-21

NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 9

LENGTH: 1753

TYPE: PRT

ORGANISM: Homo sapiens

US-10-719-385-9

Query Match 100.0%; Score 14; DB 17; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
DB 1 MIRSKITSVLSFC 14

RESULT 10  
US-10-719-385-10

Sequence 10, Application US/10719385

Publication No. US20040209284A1

GENERAL INFORMATION:

APPLICANT: O'Toole et al.

TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis

FILE REFERENCE: 22058-582

CURRENT APPLICATION NUMBER: US/10/719,385

CURRENT FILING DATE: 2003-11-21

PRIOR APPLICATION NUMBER: PCT/US03/37339

PRIOR FILING DATE: 2003-11-21

PRIOR APPLICATION NUMBER: 60/428,094

PRIOR FILING DATE: 2002-11-21

NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 10

LENGTH: 1753  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-719-385-10

Query Match 100.0%; Score 14; DB 17; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
DB 1 MIRSKITSVLSFC 14

RESULT 11  
US-10-719-385-11

Sequence 11, Application US/10719385  
Publication No. US20040209284A1  
GENERAL INFORMATION:

APPLICANT: O'Toole et al.

TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis

FILE REFERENCE: 22058-582

CURRENT APPLICATION NUMBER: US/10/719,385

CURRENT FILING DATE: 2003-11-21

PRIOR APPLICATION NUMBER: PCT/US03/37339

PRIOR FILING DATE: 2003-11-21

PRIOR APPLICATION NUMBER: 60/428,094

PRIOR FILING DATE: 2002-11-21

NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 11

LENGTH: 1753

TYPE: PRT

ORGANISM: Homo sapiens

US-10-719-385-11

Query Match 100.0%; Score 14; DB 17; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
DB 1 MIRSKITSVLSFC 14

RESULT 12  
US-10-719-385-12

Sequence 12, Application US/10719385  
Publication No. US20040209284A1  
GENERAL INFORMATION:

APPLICANT: O'Toole et al.

TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis

FILE REFERENCE: 22058-582

CURRENT APPLICATION NUMBER: US/10/719,385

CURRENT FILING DATE: 2003-11-21

PRIOR APPLICATION NUMBER: PCT/US03/37339

PRIOR FILING DATE: 2003-11-21

PRIOR APPLICATION NUMBER: 60/428,094

PRIOR FILING DATE: 2002-11-21

NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 12

LENGTH: 1753

TYPE: PRT

ORGANISM: Homo sapiens

US-10-719-385-12

Query Match 100.0%; Score 14; DB 17; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
DB 1 MIRSKITSVLSFC 14

```
Db      1 MIRSKITSVLSFC 14

RESULT 13
US-10-719-385-13
; Sequence 13, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-13

Query Match      100.0%; Score 14; DB 17; Length 1753;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MIRSKITSVLSFC 14
       |||||
Db      1 MIRSKITSVLSFC 14

RESULT 14
US-10-719-385-14
; Sequence 14, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-14

Query Match      100.0%; Score 14; DB 17; Length 1753;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MIRSKITSVLSFC 14
       |||||
Db      1 MIRSKITSVLSFC 14

RESULT 15
US-10-719-385-15
; Sequence 15, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582

; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-15

Query Match      100.0%; Score 14; DB 17; Length 1753;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MIRSKITSVLSFC 14
       |||||
Db      1 MIRSKITSVLSFC 14

RESULT 16
US-10-719-385-16
; Sequence 16, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-16

Query Match      100.0%; Score 14; DB 17; Length 1753;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MIRSKITSVLSFC 14
       |||||
Db      1 MIRSKITSVLSFC 14

RESULT 17
US-10-719-385-17
; Sequence 17, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1753
; TYPE: PRT
```

ORGANISM: Homo sapiens  
US-10-719-385-17

Query Match 100.0%; Score 14; DB 17; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
DB 1 MIRSKITSVLSFC 14

RESULT 18  
US-10-719-385-18  
Sequence 18, Application US/10719385  
Publication No. US20040209284A1

GENERAL INFORMATION:  
APPLICANT: O'Toole et al.  
TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
FILE REFERENCE: 22058-582  
CURRENT FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: PCT/US03/37339  
PRIOR FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: 60/428,094  
PRIOR FILING DATE: 2002-11-21  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 18  
LENGTH: 1753  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-719-385-18

Query Match 100.0%; Score 14; DB 17; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
DB 1 MIRSKITSVLSFC 14

RESULT 19  
US-10-719-385-19

Sequence 19, Application US/10719385  
Publication No. US20040209284A1  
GENERAL INFORMATION:  
APPLICANT: O'Toole et al.  
TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
FILE REFERENCE: 22058-582  
CURRENT FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: PCT/US03/37339  
PRIOR FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: 60/428,094  
PRIOR FILING DATE: 2002-11-21  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 19  
LENGTH: 1753  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-719-385-19

Query Match 100.0%; Score 14; DB 17; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
DB 1 MIRSKITSVLSFC 14

RESULT 20  
US-10-424-599-148727  
Sequence 148727, Application US/10424599  
Publication No. US20040031072A1

GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223) B  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 148727  
LENGTH: 165  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_105322C.1.pcp  
US-10-424-599-148727

Query Match 50.0%; Score 7; DB 15; Length 165;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSVL 11  
DB 17 SKITSVL 23

RESULT 21  
US-10-781-014-694

Sequence 694, Application US/10781014  
Publication No. US20040180408A1  
GENERAL INFORMATION:  
APPLICANT: Pompejus, Markus  
APPLICANT: Kroger, Burkhard  
APPLICANT: Schroder, Hartwig  
APPLICANT: Zelder, Oskar  
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS  
INVOLVED IN CARBON METABOLISM AND ENERGY  
FILE REFERENCE: BGI-126CPCN  
CURRENT FILING DATE: 2004-02-17  
PRIOR APPLICATION NUMBER: US/10/781,014  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 60/141,031  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: 60/143,208  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: 60/151,572  
PRIOR FILING DATE: 1999-08-31  
PRIOR APPLICATION NUMBER: DE 19931412.8  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931413.6  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931419.5  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931420.9  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931424.1  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931428.4  
PRIOR FILING DATE: 1999-07-08  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 784  
SEQ ID NO 694  
LENGTH: 285

TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-10-781-014-694

Query Match 50.0%; Score 7; DB 16; Length 285;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLSF 13  
Db 139 ITSVLSF 145

## RESULT 22

US-10-425-114-47775  
Sequence 47775, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(5313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 47775  
LENGTH: 290  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: clone ID: LIB360-022-B11\_F11.pcp  
US-10-425-114-47775

Query Match 50.0%; Score 7; DB 15; Length 290;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLSF 13  
Db 155 ITSVLSF 161

## RESULT 23

US-10-424-599-158692  
Sequence 158692, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53123)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 158692  
LENGTH: 363  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_114319C.1.pcp  
US-10-424-599-158692

Query Match 50.0%; Score 7; DB 15; Length 363;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 KITSVLS 12  
Db 135 KITSVLS 141

## RESULT 24

US-09-815-242-10735  
Sequence 10735, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zykend, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: EITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10735  
LENGTH: 431  
TYPE: PRT  
ORGANISM: Enterococcus faecalis  
US-09-815-242-10735

Query Match 50.0%; Score 7; DB 9; Length 431;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 KITSVLS 12  
Db 398 KITSVLS 404

## RESULT 25

US-10-282-122A-57038  
Sequence 57038, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zykend, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms



FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 57038  
LENGTH: 431  
TYPE: PRT  
ORGANISM: Enterococcus faecalis  
US-10-282-122A-57038

Query Match 50.0%; Score 7; DB 15; Length 431;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 KITSVLS 12  
|||  
Db 398 KITSVLS 404

RESULT 26  
US-09-738-626-6499  
Sequence 6499, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: YATEISHI, NAOKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 248-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 6499  
LENGTH: 1019  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6499

Query Match 50.0%; Score 7; DB 9; Length 1019;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 ITSLSF 13  
|||  
Db 139 ITSLSF 145

RESULT 27  
US-10-380-055-12  
Sequence 12, Application US/10380055  
Publication No. US20040014180A1  
GENERAL INFORMATION:  
APPLICANT: BOTT, Michael, et al.  
TITLE OF INVENTION: A Method For Microbial Production Of Metabolic Products, Polynucle  
FILE REFERENCE: 032301 WN 331  
CURRENT APPLICATION NUMBER: US/10/380,055  
CURRENT FILING DATE: 2003-03-11  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 12  
LENGTH: 1019  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-10-380-055-12

Query Match 50.0%; Score 7; DB 15; Length 1019;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 ITSLSF 13  
|||  
Db 139 ITSLSF 145

RESULT 28  
US-10-424-599-148728  
Sequence 148728, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 148728  
LENGTH: 1065  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_105323C.1.pcp  
US-10-424-599-148728

Query Match 50.0%; Score 7; DB 15; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 SKITSVL 11  
|||  
Db 739 SKITSVL 745

RESULT 29  
US-10-075-869-10  
Sequence 10, Application US/10075869

Publication No. US20030104622A1  
GENERAL INFORMATION:  
APPLICANT: Robbins, Paul D.  
APPLICANT: M1, Zhibao  
APPLICANT: Frizzeil, Raymond  
APPLICANT: Glorioso, Joseph C.  
APPLICANT: Gambotto, Andrea  
TITLE OF INVENTION: IDENTIFICATION OF PEPTIDES THAT  
TITLE OF INVENTION: FACILITATE UPTAKE AND CYTOPLASMIC AND/OR NUCLEAR TRANSPORT  
FILE REFERENCE: AP32573-AA 072396.0237  
CURRENT FILING DATE: 2002-02-13  
PRIOR APPLICATION NUMBER: US/10/075,869  
PRIOR FILING DATE: 1999-09-01  
PRIOR APPLICATION NUMBER: 60/151,980  
PRIOR FILING DATE: 2000-03-13  
NUMBER OF SEQ ID NOS: 99  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: random peptide library  
US-10-075-869-10

Query Match 42.9%; Score 6; DB 14; Length 12;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 IRKSKI 7  
Db 3 IRKSKI 8

RESULT 30  
US-10-366-493-10  
Sequence 10, Application US/10366493  
Publication No. US20030219826A1  
GENERAL INFORMATION:  
APPLICANT: Robbins, Paul D.  
APPLICANT: M1, Zhibao  
APPLICANT: Frizzeil, Raymond  
APPLICANT: Glorioso, Joseph C.  
APPLICANT: Gambotto, Andrea  
TITLE OF INVENTION: IDENTIFICATION OF PEPTIDES THAT FACILITATE UPTAKE AND CYTOPLASMIC  
TITLE OF INVENTION: OF PROTEINS, DNA AND VIRUSES  
FILE REFERENCE: AP32573-A-A-A-A 072396.0246  
CURRENT APPLICATION NUMBER: US/10/366,493  
CURRENT FILING DATE: 2003-02-12  
PRIOR APPLICATION NUMBER: 10/075,869  
PRIOR FILING DATE: 2002-02-13  
PRIOR APPLICATION NUMBER: 09/653,182  
PRIOR FILING DATE: 2000-08-31  
PRIOR APPLICATION NUMBER: 60/188,944  
PRIOR FILING DATE: 2000-03-13  
PRIOR APPLICATION NUMBER: 60/151,980  
PRIOR FILING DATE: 1999-09-01  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: pep 10  
US-10-366-493-10

Query Match 42.9%; Score 6; DB 14; Length 12;  
Best Local Similarity 100.0%; Pred. No. 23;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 IRKSKI 7  
Db 3 IRKSKI 8

RESULT 31  
US-10-029-347-10  
Sequence 10, Application US/10029347  
Publication No. US20030017562A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: A NOVEL HUMAN LEUCINE-RICH REPEAT CONTAINING PROTEIN EXPRESSED  
TITLE OF INVENTION: PREDOMINATELY IN SMALL INTESTINE, HLRSL1  
FILE REFERENCE: D0066  
CURRENT APPLICATION NUMBER: US/10/029,347  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 10  
LENGTH: 17  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-029-347-10

Query Match 42.9%; Score 6; DB 14; Length 17;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ITSVLS 12  
Db 7 ITSVLS 12

RESULT 32  
US-10-437-963-111764  
Sequence 111764, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Bouharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 111764  
LENGTH: 48  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_15711C.1.pep  
US-10-437-963-111764

Query Match 42.9%; Score 6; DB 16; Length 48;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 SVLSFC 14  
Db 40 SVLSFC 45

RESULT 33  
US-10-424-599-267615

; Sequence 267615, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 267615  
; LENGTH: 53  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_8367C.1.pep  
US-10-424-599-267615

Query Match 42.9%; Score 6; DB 15; Length 53;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
DB 3 ITSVLS 8

RESULT 34  
US-10-425-115-185349  
; Sequence 185349, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 185349  
; LENGTH: 56  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_100626C.1.pep  
US-10-425-115-185349

Query Match 42.9%; Score 6; DB 17; Length 56;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKITS 9  
DB 47 KSKITS 52

RESULT 35  
US-10-437-963-197926  
; Sequence 197926, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 197926  
; LENGTH: 60  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_93637C.1.pep  
US-10-437-963-197926

Query Match 42.9%; Score 6; DB 16; Length 60;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKITS 9  
DB 53 KSKITS 58

RESULT 36  
US-10-424-599-284793  
; Sequence 284793, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 284793  
; LENGTH: 69  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_99196C.1.pep  
US-10-424-599-284793

Query Match 42.9%; Score 6; DB 15; Length 69;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
DB 8 SKITSV 13

RESULT 37  
US-10-424-599-143794  
; Sequence 143794, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 143794

LENGTH: 82  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_10085C.1.pep  
US-10-424-599-143794

Query Match 42.9%; Score 6; DB 15; Length 82;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IRKSKI 7  
19 IRKSKI 24

RESULT 38  
US-10-424-599-217068  
Sequence 217068, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 217068  
LENGTH: 100  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_3803C.1.pep  
US-10-424-599-217068

Query Match 42.9%; Score 6; DB 15; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSUUSF 13  
73 TSUUSF 78

RESULT 39  
US-10-425-115-307173  
Sequence 307173, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 307173  
LENGTH: 100  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_43212C.1.pep  
US-10-425-115-307173

Query Match 42.9%; Score 6; DB 17; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 5 SKITSV 10  
48 SKITSV 53

RESULT 40  
US-10-425-114-54103  
Sequence 54103, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(5313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 54103  
LENGTH: 111  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: 700577676\_F11.pep  
US-10-425-114-54103

Query Match 42.9%; Score 6; DB 15; Length 111;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 KITSVL 11  
15 KITSVL 20

Search completed: December 30, 2004, 15:48:49  
Job time : 147 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 30, 2004, 14:45:25 ; Search time 0.934622 Seconds  
(without alignment)  
3693.737 Million cell updates/sec

Title: US-10-719-385-26

Perfect score: 36

Sequence: 1 AGGRCV 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Uniprot\_02:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	36	100.0	339	2 Q7XS00	Q7XS00 oryza sativ
2	36	100.0	407	2 Q9UPF4	Q9UPF4 homo sapien
3	36	100.0	444	1 Y966_METTA	Y966_METTA
4	36	100.0	469	2 O28589	O28589 archaeglob
5	36	100.0	499	2 Q7SYVO	Q7SYVO xenopus lae
6	36	100.0	798	2 Q8C7A1	Q8C7A1 mus musculu
7	36	100.0	843	2 Q6GQ09	Q6GQ09 homo sapien
8	36	100.0	858	2 Q9N0S3	Q9N0S3 homo sapien
9	36	100.0	1639	2 Q7Z3K8	Q7Z3K8 homo sapien
10	36	100.0	1745	2 Q14675	Q14675 homo sapien
11	36	100.0	1761	2 Q6Z0H8	Q6Z0H8 mus musculu
12	36	100.0	1761	2 BAC97883	BAC97883 mus muscu
13	35	97.2	324	2 O8S345	O8S345 capsicum an
14	35	97.2	383	2 Q50225	Q50225 methanosarc
15	35	97.2	383	2 Q50248	Q50248 methanosarc
16	35	97.2	386	2 Q8PUW0	Q8PUW0 methanosarc
17	35	97.2	386	2 Q8PUV6	Q8PUV6 methanosarc
18	35	97.2	393	2 Q8A8W8	Q8A8W8 bacteroides
19	35	97.2	398	2 Q9W505	Q9W505 vicia vilif
20	35	97.2	409	2 Q93XU1	Q93XU1 salix gligi
21	35	97.2	418	2 Q6U7H8	Q6U7H8 malus domes
22	35	97.2	418	2 AAQ84042	AAQ84042 malus dom
23	35	97.2	501	2 Q27917	Q27917 methanobac
24	35	97.2	1036	1 ACR1_HUMAN	ACR1_HUMAN
25	33	91.7	93	2 O84SX1	O84SX1 oryza sativ
26	33	91.7	94	2 O8LH84	O8LH84 oryza sativ
27	33	91.7	591	1 U149_EBV	U149_EBV
28	33	91.7	591	2 Q777G6	Q777G6 human herpe
29	33	91.7	591	2 CAD53400	CAD53400 human her
30	33	91.7	2434	2 Q6RKJ9	Q6RKJ9 botrytis ci
31	33	91.7	2434	2 AAR90237	AAR90237 botrytis

32	32	88.9	73	2 O66009	O66009 bacillus su
33	32	88.9	77	2 Q75WH6	Q75WH6 macrochele
34	32	88.9	77	2 BAD13402	BAD13402 macrochele
35	32	88.9	91	2 Q7UYG4	Q7UYG4 rhodospirilli
36	32	88.9	106	2 Q9CTQ0	Q9CTQ0 mus musculu
37	32	88.9	108	2 Q8BPN6	Q8BPN6 mus musculu
38	32	88.9	119	2 Q7KUM2	Q7KUM2 drosophila
39	32	88.9	119	2 AAS65002	AAS65002 drosophila
40	32	88.9	132	2 O6SHL1	O6SHL1 uncultured
41	32	88.9	132	2 AAR37610	AAR37610 uncultured
42	32	88.9	136	2 Q25300	Q25300 leishmania
43	32	88.9	137	2 Q6G3N4	Q6G3N4 bartonella
44	32	88.9	140	2 Q87XH2	Q87XH2 pseudomonas
45	32	88.9	147	2 Q7WY76	Q7WY76 bordetella
46	32	88.9	148	2 Q8LR19	Q8LR19 oryza sativ
47	32	88.9	156	2 Q7ZB17	Q7ZB17 desulfovibrio
48	32	88.9	156	2 AAS95242	AAS95242 desulfovibrio
49	32	88.9	158	2 Q943U7	Q943U7 oryza sativ
50	32	88.9	200	2 Q95HV9	Q95HV9 peromyscus
51	32	88.9	200	2 Q95HW1	Q95HW1 peromyscus
52	32	88.9	200	2 Q9RUI2	Q9RUI2 streptomyces
53	32	88.9	206	2 Q6G3N6	Q6G3N6 bartonella
54	32	88.9	208	2 Q94243	Q94243 schizosacch
55	32	88.9	214	2 Q9F6W0	Q9F6W0 rhizobium e
56	32	88.9	216	2 Q8N7M1	Q8N7M1 homo sapien
57	32	88.9	218	2 Q96NZ0	Q96NZ0 homo sapien
58	32	88.9	229	2 Q82EH8	Q82EH8 streptomyces
59	32	88.9	247	2 Q7T9U2	Q7T9U2 adoxophyes
60	32	88.9	253	2 Q8N8T0	Q8N8T0 homo sapien
61	32	88.9	254	2 Q81604	Q81604 mesembryant
62	32	88.9	256	2 P72460	P72460 streptomyces
63	32	88.9	266	2 Q8NAJ1	Q8NAJ1 homo sapien
64	32	88.9	269	2 Q7X6S7	Q7X6S7 oryza sativ
65	32	88.9	269	2 CAD41855	CAD41855 oryza sat
66	32	88.9	269	2 CAE05968	CAE05968 oryza sat
67	32	88.9	275	2 Q9ALM7	Q9ALM7 saccharopol
68	32	88.9	280	2 P96821	P96821 mycobacteri
69	32	88.9	280	2 Q7U2R4	Q7U2R4 mycobacteri
70	32	88.9	287	2 Q7XK68	Q7XK68 oryza sativ
71	32	88.9	292	1 PER2_CUCSA	PER2_CUCSA
72	32	88.9	312	2 Q8L442	Q8L442 oryza sativ
73	32	88.9	314	2 Q9NE71	Q9NE71 leishmania
74	32	88.9	320	2 Q982F1	Q982F1 rhizobium l
75	32	88.9	322	2 Q42964	Q42964 nicotiana t
76	32	88.9	323	2 Q6K770	Q6K770 oryza sativ
77	32	88.9	323	2 BAD21872	BAD21872 oryza sat
78	32	88.9	323	2 BAD21877	BAD21877 oryza sat
79	32	88.9	324	2 Q90XK3	Q90XK3 gnathomonas
80	32	88.9	326	2 Q90XK3	Q90XK3 gnathomonas
81	32	88.9	341	2 Q98VG3	Q98VG3 human herpe
82	32	88.9	342	2 Q6Z135	Q6Z135 oryza sativ
83	32	88.9	342	2 BAD15483	BAD15483 oryza sat
84	32	88.9	352	2 Q82MX5	Q82MX5 streptomyces
85	32	88.9	374	1 SBP_CRYTA	SBP_CRYTA
86	32	88.9	374	2 Q8RUI1	Q8RUI1 cryptomeria
87	32	88.9	382	2 Q8TUD9	Q8TUD9 methanosarc
88	32	88.9	398	2 Q43783	Q43783 musa acumin
89	32	88.9	398	2 Q7V908	Q7V908 prochloroc
90	32	88.9	398	2 Q93OR4	Q93OR4 rhizobium m
91	32	88.9	404	1 PE18_LYCES	PE18_LYCES
92	32	88.9	406	2 Q8T3W6	Q8T3W6 drosophila
93	32	88.9	407	1 EXK1_HUMAN	EXK1_HUMAN
94	32	88.9	407	2 Q9SDW4	Q9SDW4 musa acumin
95	32	88.9	408	1 P118_ARATH	P118_ARATH
96	32	88.9	423	2 Q93256	Q93256 gallus galli
97	32	88.9	425	1 Y450_HUMAN	Y450_HUMAN
98	32	88.9	426	2 Q97KZ7	Q97KZ7 clostridium
99	32	88.9	432	1 P122_ARATH	P122_ARATH
100	32	88.9	437	1 E2F1_HUMAN	E2F1_HUMAN
101	32	88.9	438	2 Q6L168	Q6L168 methanococ
102	32	88.9	438	2 CAF30679	CAF30679 methanococ
103	32	88.9	448	1 FXN4_HUMAN	FXN4_HUMAN
104	32	88.9	450	1 VD10_BPT5	VD10_BPT5 bacterioph

105 32 88.9 450 2 AAS7170 bacteriop  
106 32 88.9 453 2 O892A3 clostridium  
107 32 88.9 454 2 O969U8 clostridium  
108 32 88.9 454 2 O97GT6 clostridium  
109 32 88.9 455 2 O8ECH2 shewanella  
110 32 88.9 455 2 O9ET11 mus musculus  
111 32 88.9 455 2 O920R1 mus musculus  
112 32 88.9 457 2 O8XJ18 mus musculus  
113 32 88.9 462 1 SYG\_TREPA  
114 32 88.9 462 2 O9HD26  
115 32 88.9 463 2 O8BH60  
116 32 88.9 478 2 O70PA5  
117 32 88.9 478 2 O6N9V7  
118 32 88.9 478 2 CAD89766  
119 32 88.9 478 2 CAE26872  
120 32 88.9 479 1 FYLG\_HUMAN  
121 32 88.9 497 2 P90338  
122 32 88.9 503 2 O6IEL5  
123 32 88.9 506 2 O6Z8E9  
124 32 88.9 506 2 BAD07870  
125 32 88.9 508 2 O8K234  
126 32 88.9 508 2 O9D682  
127 32 88.9 513 2 O7PIU8  
128 32 88.9 531 1 GHT2\_SCHPO  
129 32 88.9 534 2 GHT6\_SCHPO  
130 32 88.9 587 2 O88831  
131 32 88.9 588 2 O8C078  
132 32 88.9 588 2 O8CH42  
133 32 88.9 588 2 O8CH42  
134 32 88.9 616 2 O8RBB3  
135 32 88.9 617 2 O8X116  
136 32 88.9 617 2 O94917  
137 32 88.9 672 2 O7US07  
138 32 88.9 687 2 O9XMD1  
139 32 88.9 699 1 VGLG\_HHV2H  
140 32 88.9 717 2 O922U5  
141 32 88.9 719 2 O82MW5  
142 32 88.9 744 2 O12715  
143 32 88.9 747 2 O6UYU0  
144 32 88.9 747 2 AAO76093  
145 32 88.9 750 2 O6Z1X0  
146 32 88.9 750 2 BAC92820  
147 32 88.9 755 2 P89107  
148 32 88.9 773 2 O6UTU1  
149 32 88.9 773 2 AAO56304  
150 32 88.9 791 2 O96L64

## ALIGNMENTS

RESULT 1  
Q7XS00 PRELIMINARY; PRT; 339 AA.

AC Q7XS00, 01-OCT-2003 (TRENBLrel. 25, Created)  
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE OSJNB0084K11.19 protein.  
GN Name=OSJNB0084K11.19;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=33947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=12447439;  
RA Liu Y., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,  
RA Feng Q., Zhang Y., Jia P., Zhang Y., Zhao Q., Yang K., Yu S., Tang Y.,  
RA Weng Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L.S., Yu Z., Fan D.,  
RA Liu X., Li T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,  
RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,

RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,  
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,  
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,  
RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,  
RA Han B.,  
RT "Sequence and analysis of rice chromosome 4."  
RL Nature 420:316-320(2002).  
DR EMBL; AL606687; CAE01851.2; -  
DR Gramene; O7XS00; -  
DR GO; GO:0046872; F-metal ion binding; IEA.  
DR InterPro; IPR001594; Znf\_DHHC.  
DR Pfam; PF01529; zf-DHHC; 1.  
DR ProDom; PD003041; Znf\_DHHC; 1.  
DR PROSITE; PS50216; ZF\_DHHC; 1.  
SQ SEQUENCE 339 AA; 3896 MW; 56F6B67F2A2650E CRC64;

Query Match 100.0%; Score 36; DB 2; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6  
DB 6 AGGPCV 11

RESULT 2  
Q9UFP4 PRELIMINARY; PRT; 407 AA.

AC Q9UFP4, 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE Hypothetical protein DKFZp344H0717 (Fragment).  
GN Name=DKFZp344H0717;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP TISSUE=Testis;  
RC Bioecker H., Boecker W., Brandt P., Mewes H.W., Gassenhuber J.,  
RA Wiemann S.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL122102; CAB59268.1; -  
DR PIR; T34535; T34535.  
KW Hypothetical protein.  
FT NON\_TER  
SQ SEQUENCE 407 AA; 43799 MW; FD4F43CE2AC0DB3C CRC64;

Query Match 100.0%; Score 36; DB 2; Length 407;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6  
DB 280 AGGPCV 285

## RESULT 3

Y966\_METUA STANDARD; PRT; 444 AA.  
AC O58376;  
DT 29-MAR-2004 (Rel. 43, Created)  
DT 29-MAR-2004 (Rel. 43, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Hypothetical protein MJ0966.  
GN OrderedLocustNames=MJ0966;  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=21907;  
RN [1]



```

RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Overbeek R., Kirkness E.F., Weissbrock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Feldman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073 (1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, U67539; AAB89868.1; -.
CC PIR, F64420; F64420.
CC TIGR, MJ0966; -.
CC InterPro, IPR006638; E1p3/M1AB/N1fB.
CC InterPro, IPR007197; Radical SAM.
CC Pfam, PF04055; Radical SAM; I.
CC SMART, SM00729; E1p3; I.
CC Complete proteome; Hypothetical protein.
CC SEQUENCE 444 AA; 51347 MW; 25ABAB5890A20C70 CRC64;

Query Match 100.0%; Score 36; DB 1; Length 444;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGPCV 6
Db 88 AGPCV 93

RESULT 4
028589 PRELIMINARY; PRT; 469 AA.
AC 028589;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein AFI684.
GN OrderedLocusNames=AFI684;
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=96049343; PubMed=9389475; DOI=10.1038/37052;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Kerechum K.A., Dodson R.J., Gwinn M.L., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.R.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.J.,
RA Peterson S.N., Reich C.I., McNeil L.R., Badger J.H., Glodek A.,
RA Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L.A.,
RA Uterback T.R., Cotton M.D., Spriggs T., Artiach P., Kaine B.P.,
RA Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C.,
RA Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O.,
RA Woese C.R., Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370 (1997).

```

```

DR EMBL, AE000987; AAB89565.1; -.
DR PIR, C69460; C69460.
DR TIGR, AF1684; -.
DR InterPro, IPR006638; E1p3/M1AB/N1fB.
DR InterPro, IPR007197; Radical SAM.
DR Pfam, PF04055; Radical SAM; I.
DR SMART, SM00729; E1p3; I.
CC Complete proteome; Hypothetical protein.
CC SEQUENCE 469 AA; 53468 MW; 0FE86155B1P1004 CRC64;

Query Match 100.0%; Score 36; DB 2; Length 469;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGPCV 6
Db 104 AGPCV 109

RESULT 5
0751V0 PRELIMINARY; PRT; 499 AA.
AC 0751V0;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Parsl-prov protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Alechul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Heien F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heltan E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalins D.E., Schermer A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RX [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT Initiative."
RL Dev. Dyn. 225:384-391 (2002).
RX [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RA Klein S., Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
CC -1- Similarity: Belongs to class-II aminoacyl-tRNA synthetase family.
DR EMBL, BC054256; AAB44256.1; -.
DR GO, GO:0005737; C:cytoplasm; IEA.

```

```

DR GO:0005524; F:ATP binding; IEA.
DR GO:0004826; F:phenylalanine-tRNA ligase activity; IEA.
DR GO:0006432; F:phenylalanine-tRNA aminoacylation; IEA.
DR InterPro: IPR004529; Pfam.
DR InterPro: IPR002319; tRNA-synt_2d.
DR InterPro: IPR006195; tRNA_ligase_II.
DR Pfam: PF01409; tRNA-synt_3d; 1.
DR TIGRFAMs: TIGR00468; pbes; 1.
DR PROSITE: PS50862; AA TRNA_LIGASE_II; 1.
DR SEQUENCE 499 AA; 56750 MW; A632A680B2C27C2 CRC64;

Query Match 100.0%; Score 36; DB 2; Length 499;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGPCV 6
Db 121 AGPCV 126

RESULT 6
O8C7A1 PRELIMINARY; PRT; 798 AA.
AC O8C7A1;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched
DE library, clone:D330017D22 product:hypothetical protein, full insert
DE sequence.
GN Name=BC025526;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
ON (1)
RN R1 SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Mech. Enzymol. 303:19-44(1999).
[2]
RN R2 SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RN R3 SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RA The FANTOM Consortium;
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
[4]
RN R4 SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=2049374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RN R5 SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Kikunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,

```

```

RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaishwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Matsuki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN R6 SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RA Adachi U., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirokane T.,
RA Hoti F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai U., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ono M., Ohnato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK052272; BAC34911.1; -
DR MGD; MGI:2446190; BC025526.
KW Hypothetical protein.
SQ SEQUENCE 798 AA; 90029 MW; BA333ABB589BAFAC CRC64;

Query Match 100.0%; Score 36; DB 2; Length 798;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGPCV 6
Db 5 AGPCV 10

RESULT 7
O6G009 PRELIMINARY; PRT; 843 AA.
ID O6G009;
AC O6G009;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE ZAZ0D1 protein.
GN Name=ZAZ0D1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
ON (1)
RN R1 SEQUENCE FROM N.A.
RC TISSUE=Bye;
RX MEDLINE=22388257; PubMed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer J., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Diatchenko L., Marisina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein W.J., Udell T.B., Yoshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McGwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettunen M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield J.S.,
RA Kravinsky M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,
RA Jones S.D., Warr M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strauberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC072681; AA072681.1; -.
DR InterPro; IPR003323; OTU.
DR InterPro; IPR002653; Znf_A20.
DR Pfam; PF02338; OTU; 1.
DR SMART; SM00259; Znf_A20; 1.
DR PROSITE; PSS0802; OTU; 1.
SQ SEQUENCE 843 AA; 92525 MW; 6D386C864B12EE57 CRC64;

Query Match
Best Local Similarity 100.0%; Score 36; DB 2; Length 843;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6
DB 716 AGGPCV 721

RESULT 8
Q9NO53 PRELIMINARY; PRT; 858 AA.
ID Q9NO53
AC Q9NO53;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Zinc finger protein Cezanne.
GN Name=CEZANNE;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21356309; PubMed=11463333;
RA Evans P.C., Taylor E.R., Coadwell J., Heynck K., Beyaert R.,
RA Klishaw F.J.;
RT "Isolation and characterization of two novel A20-like proteins.";
RL Biochem. J. 357:617-623(2001).
DR EMBL; AJ293573; CAB97494.1; -.
DR MEROPS; C64.001; -.
DR Genew; HGNC:16683; ZA20D1.
DR InterPro; IPR003323; OTU.
DR InterPro; IPR009060; UBA_1like.
DR Pfam; PF02338; OTU; 1.
DR PROSITE; PSS0802; OTU; 1.
SQ SEQUENCE 858 AA; 94400 MW; A7D6BD280C9387F CRC64;

Query Match
Best Local Similarity 100.0%; Score 36; DB 2; Length 858;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6
DB 731 AGGPCV 736

RESULT 9
Q723K8 PRELIMINARY; PRT; 1639 AA.
ID Q723K8
AC Q723K8;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein DKFZp686l1653 (Fragment).
GN Name=DKFZp686l1653;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human endometrium;
RA Wandut R., Heudner D., Mewes H.W., Weil B., Amid C., Oanger A.,
RA Pobo G., Han M., Wiemann S.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX537774; CAD97835.1; -.
DR Hypothetical protein.
FT NON_TER
SQ SEQUENCE 1639 AA; 182478 MW; B58334E1B217B92 CRC64;

Query Match
Best Local Similarity 100.0%; Score 36; DB 2; Length 1639;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6
DB 6 AGGPCV 11

RESULT 10
Q14675 PRELIMINARY; PRT; 1745 AA.
ID Q14675
AC Q14675;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE KIA0169 protein (Fragment).
GN Name=KIA0169;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIA0161-KIA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 3:17-24(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Ohara O., Nagase T., Kikuno R., Nomura N.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; D79991; BA011486.1; -.
DR Genew; HGNC:17859; NUP188.
FT NON_TER
FT NON_TER
SQ SEQUENCE 1745 AA; 195637 MW; CF74DBA9CF6B871 CRC64;

Query Match
Best Local Similarity 100.0%; Score 36; DB 2; Length 1745;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6
DB 1 AGGPCV 6

RESULT 11
Q6ZOH8 PRELIMINARY; PRT; 1761 AA.
ID Q6ZOH8
AC Q6ZOH8;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE MKIA0169 protein (Fragment).
GN Name=MKIA0169;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```

```

RN SEQUENCE FROM N.A.
RP TISSUE=Embryonic tail;
RC PubMed=14621295;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
Saga Y., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIA gene:
RT ii. the complete nucleotide sequences of 500 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:167-180(2003).
DR EMBL; AK129073; BAC97883.1; -.
DR InterPro; IPR008938; ARM.
DR NON TER 1 1
SQ SEQUENCE 1761 AA; 196893 MW; 20CE3DC250EBD93 CRC64;

Query Match 100.0%; Score 36; DB 2; Length 1761;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGRCV 6
Db 7 AGGRCV 12

RESULT 12
BAC97883 PRELIMINARY; PRT; 1761 AA.
AC BAC97883;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE MKIA0169 protein (Fragment).
GN MKIA0169.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic tail;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
Saga Y., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of Mouse Homologues of KIA Gene:
RT ii. The complete nucleotide sequences of 500 Mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:167-180(2003).
DR EMBL; AK129073; BAC97883.1; -.
DR NON TER 1 1
SQ SEQUENCE 1761 AA; 196893 MW; 20CE3DC250EBD93 CRC64;

Query Match 100.0%; Score 36; DB 2; Length 1761;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGRCV 6
Db 7 AGGRCV 12

RESULT 13
OBS345 PRELIMINARY; PRT; 324 AA.
AC OBS345;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative pectate-lyase (Fragment).
GN Name=upae;
OS Capsicum annuum (Bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

```

```

OC Lamiids; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxId=4072;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22113994; PubMed=12116879;
RA Marcis E., Van den Ackerveken G., Bonas U.;
RT "The xanthomonas type III effector protein AvrBs3 modulates plant gene
RT expression and induces cell hypertrophy in the susceptible host.";
RL Mol. Plant Microbe Interact. 15:637-646(2002).
DR EMBL; AF492632; AAM12784.1; -.
DR GO; GO:0016829; F:lyase activity; IEA.
DR InterPro; IPR002022; Amb_allergen.
DR InterPro; IPR006626; PbH1.
DR InterPro; IPR01050; Pectin_lyas_1like.
DR Pfam; PF00544; Pect_lyase_C_1.
DR PRINTS; PRO0807; AMBALERGEN.
DR SMART; SM00656; Amb_all; 1.
DR SMART; SM00710; PbH1; 2.
DR LYASE.
DR NON TER 324 324
SQ SEQUENCE 324 AA; 35751 MW; BC2BED36F8B96B7A CRC64;

Query Match 97.2%; Score 35; DB 2; Length 324;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGRCV 6
Db 155 AGGRCV 160

RESULT 14
O50225 PRELIMINARY; PRT; 383 AA.
AC O50225;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE F420-nonreducing hydrogenase (Small subunit (40kDa) of the membrane-
DE bound hydrogenase ii) precursor (EC 1.12.99.-).
GN Name=whc;
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxId=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC SMPAIN=GO1;
RX MEDLINE=95154297; PubMed=7851393;
RA Deppenmeier U., Blaut M., Lenters S., Herzberg C., Gottschalk G.;
RT "Analysis of the vhoGAC and vhoAC operons from Methanosarcina mazei
RT strain Go1 both encoding a membrane-bound hydrogenase and a cytochrome
RT b " ".
RL Eur. J. Biochem. 227:261-269(1995).
DR EMBL; X83112; CA58176.1; -.
DR PIR; S67477; S67477.
DR HSSP; P13063; ICCL.
DR GO; GO:0009375; C:hydrogenase complex; IEA.
DR GO; GO:0008901; F:ferredoxin hydrogenase activity; IEA.
DR GO; GO:000506; F:iron ion binding; IEA.
DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO; GO:0016481; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
DR InterPro; IPR001821; NlFenydrog_small.
DR InterPro; IPR006317; Oxidored_q6.
DR Pfam; PF01058; Oxidored_q6; 1.
DR TIGRFAMs; TIGR00391; hycA; 1.
DR TIGRFAMs; TIGR01409; hAT_signal_seq; 1.
DR Oxidoreductase; Signal.
KW SIGNAL.
FT CHAIN 1 48 Potential.
FT 49 383 Potential.

```

SQ SEQUENCE 383 AA; 41152 MW; 0F35036868141456 CRC64;  
 Query Match 97.2%; Score 35; DB 2; Length 383;  
 Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPGCV 6  
 Db 323 AGGPCT 328  
 RESULT 15  
 Q50248 PRELIMINARY; PRT; 383 AA.  
 ID 050248  
 AC 050248  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE F420-nonreducing hydrogenase I precursor.  
 GN Name=vhog;  
 OS Methanosarcina mazel (Methanosarcina frisia).  
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2209;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Goel;  
 RX MEDLINE=95154297; PubMed=7851393;  
 RA Deppenmeier U., Blaut M., Lentes S., Herzberg C., Gottschalk G.;  
 RT "Analysis of the vhoGAC and vhtGAC operons from Methanosarcina mazel  
 RT strain Goel both encoding a membrane-bound hydrogenase and a cytochrome  
 b";  
 RT Eur. J. Biochem. 227:261-269(1995).  
 DR EMBL; X82940; CAA58113.1; -.  
 DR FIR; S49928; S49928.  
 DR HSSP; P13063; ICCL.  
 DR GO; GO:0009375; C:hydrogenase complex; IEA.  
 DR GO; GO:0008901; F:ferredoxin hydrogenase activity; IEA.  
 DR GO; GO:0005506; F:iron ion binding; IEA.  
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.  
 DR InterPro; IPR001821; Nifendrog\_small.  
 DR InterPro; IPR006137; Oxidored\_g6.  
 DR Pfam; PF01058; Oxidored\_g6; 1.  
 DR TIGRFAMs; TIGR00391; hvdA; 1.  
 DR TIGRFAMs; TIGR01409; TAT\_signal\_seq; 1.  
 KW Signal.  
 FT SIGNAL. 1 48 Potential.  
 SQ SEQUENCE 383 AA; 41067 MW; 08C94F527771C9B8 CRC64;  
 Query Match 97.2%; Score 35; DB 2; Length 383;  
 Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPGCV 6  
 Db 323 AGGPCT 328  
 RESULT 16  
 Q8PUM0 PRELIMINARY; PRT; 386 AA.  
 ID Q8PUM0  
 AC Q8PUM0  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE F420-nonreducing hydrogenase I (EC 1.12.2.-).  
 GN Name=vhog; OrderedLocuNames=MM214;  
 OS Methanosarcina mazel (Methanosarcina frisia).  
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.

OX NCBI\_TaxID=2209;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Goel / Goel / ATCC BAA-199 / DSM 3647 / OCM 88;  
 RX MEDLINE=22120827; PubMed=12125824;  
 RA Deppenmeier U., Johann A., Hartesch T., Merkl R., Schmitz R.A.,  
 RA Martinez-Arias R., Henne A., Wierer A., Baeumer S., Jacobi C.,  
 RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,  
 RA Bhattacharya A., Lykilds A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
 RA Fritz H.-J., Gottschalk G.;  
 RT "The genome of Methanosarcina mazel: evidence for lateral gene  
 RT transfer between Bacteria and Archaea.";  
 RT J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
 DR EMBL; AB013474; AAM32010.1; -.  
 DR HSSP; P13063; ICCL.  
 DR GO; GO:0009375; C:hydrogenase complex; IEA.  
 DR GO; GO:0008901; F:ferredoxin hydrogenase activity; IEA.  
 DR GO; GO:0005506; F:iron ion binding; IEA.  
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.  
 DR InterPro; IPR001821; Nifendrog\_small.  
 DR InterPro; IPR006137; Oxidored\_g6.  
 DR Pfam; PF01058; Oxidored\_g6; 1.  
 DR TIGRFAMs; TIGR00391; hvdA; 1.  
 DR TIGRFAMs; TIGR01409; TAT\_signal\_seq; 1.  
 KW Complete proteome; Oxidoreductase.  
 SQ SEQUENCE 386 AA; 41427 MW; E80AA4F3B9CEFC7C CRC64;  
 Query Match 97.2%; Score 35; DB 2; Length 386;  
 Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPGCV 6  
 Db 326 AGGPCT 331  
 RESULT 17  
 Q8PVO6 PRELIMINARY; PRT; 386 AA.  
 ID Q8PVO6  
 AC Q8PVO6  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE F420-nonreducing hydrogenase II (EC 1.12.2.-).  
 GN Name=vhog; OrderedLocuNames=MM2169;  
 OS Methanosarcina mazel (Methanosarcina frisia).  
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2209;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Goel / Goel / ATCC BAA-199 / DSM 3647 / OCM 88;  
 RT J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
 DR EMBL; AB013457; AAM31865.1; -.  
 DR HSSP; P13063; ICCL.  
 DR GO; GO:0009375; C:hydrogenase complex; IEA.  
 DR GO; GO:0008901; F:ferredoxin hydrogenase activity; IEA.  
 DR GO; GO:0005506; F:iron ion binding; IEA.  
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.

```

DR GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
DR InterPro: IPR001821; NIFHydrog_small.
DR InterPro: IPR006137; Oxidored_g6.
DR InterPro: IPR006311; Tat.
DR Pfam: PF01058; Oxidored_g6; 1.
DR TIGRFAMs: TIGR00391; hYdA; 1.
DR TIGRFAMs: TIGR01409; TAT_signal_seq; 1.
DR Complete proteome; Oxidoreductase.
SQ SEQUENCE 386 AA; 41511 MW; EFFF68C9A68923C2 CRC64;

Query Match          97.2%; Score 35; DB 2; Length 386;
Best Local Similarity 83.3%; Pred. No. 2.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPGCV 6
DB 326 AGGPCT 331

RESULT 18
Q8A8W8 PRELIMINARY; PRT; 393 AA.
AC Q8A8W8:
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Putative patatin-like protein.
GN OrderedLocusNames=BT1049;
OS Bacteroides thetaiotaomicron.
OC Bacterioidetes; Bacteroidetes; Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RA MEDLINE=22550858; PubMed=12663928;
RX Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
  Chang H.C., Hooper L.V., Gordon J.I.;
  "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
  Science 299:2074-2076 (2003).
RL EMBL; AE016930; AAO76156.1; -
DR InterPro: IPR008985; Cona_like_g1.
KW Complete proteome.
SQ SEQUENCE 393 AA; 43221 MW; 9158D50A717369A CRC64;

Query Match          97.2%; Score 35; DB 2; Length 393;
Best Local Similarity 83.3%; Pred. No. 2.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPGCV 6
DB 293 AGGPCT 298

RESULT 19
Q9M5O5 PRELIMINARY; PRT; 398 AA.
AC Q9M5O5:
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Pectate lyase.
GN Name=Pel;
OS Vitis vinifera (Grape).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; rosids;
OC Vitaceae; Vitis.
OX NCBI_TaxID=29760;
RN (1)
RP SEQUENCE FROM N.A.
RA Papadakis G.N., Kanelis A.K.;
  Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
  EMBL; AF243475; AAF63756.1; -

```

```

DR HSSP; P29155; IJRG.
DR GO:0016829; F:lyase activity; IEA.
DR InterPro: IPR002022; Amb_allergen.
DR InterPro: IPR006626; Pbh1.
DR InterPro: IPR01050; Pectin_lyas_like.
DR Pfam: PF00544; Pec_lyase_C; 1.
DR PRINTS: PR00807; AMBALLERGEN.
DR SMART: SM00656; Amb_all; 1.
DR SMART: SM00710; Pbh1; 2.
SQ SEQUENCE 398 AA; 44000 MW; D20A568A1F1241F4 CRC64;

Query Match          97.2%; Score 35; DB 2; Length 398;
Best Local Similarity 83.3%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPGCV 6
DB 151 AGGPCT 156

RESULT 20
Q93XJ1 PRELIMINARY; PRT; 409 AA.
AC Q93XJ1:
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Pectate lyase.
GN Name=sgpL;
OS Salix gligiana.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC Eurosid1; Malpighiales; Salicaceae; Salix.
OX NCBI_TaxID=75556;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Calkin;
RA Futamura N., Kouchi H., Shinohara K.;
  "A gene for pectate lyase expressed in elongating and differentiating
  tissues of a Japanese willow (Salix gligiana).";
  J. Plant Physiol. 159:1123-1130(2002).
RL EMBL; AB048260; BAB59066.1; -
DR HSSP; P29155; IJRG.
DR GO:0016829; F:lyase activity; IEA.
DR InterPro: IPR002022; Amb_allergen.
DR InterPro: IPR006626; Pbh1.
DR InterPro: IPR01050; Pectin_lyas_like.
DR Pfam: PF00544; Pec_lyase_C; 1.
DR PRINTS: PR00807; AMBALLERGEN.
DR SMART: SM00656; Amb_all; 1.
DR SMART: SM00710; Pbh1; 2.
SQ SEQUENCE 409 AA; 44694 MW; F97E9D0A28CEDFCD CRC64;

Query Match          97.2%; Score 35; DB 2; Length 409;
Best Local Similarity 83.3%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPGCV 6
DB 162 AGGPCT 167

RESULT 21
Q6U7H9 PRELIMINARY; PRT; 418 AA.
AC Q6U7H9:
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Pectate lyase.
OS Malus domestica (Apple) (Malus sylvestris).

```

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucotsids I; Rosales; Rosaceae; Maloideae; Malus.
OX NCB1_TaxID=3750;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Ripe fruit;
RA Goulae L.F., Oliveira C.M.;
RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY376878; AAC84042.1; -
DR GO: GO:0016829; P:lyase activity; IEA.
DR InterPro: IPR010502; Amb_allergen.
DR InterPro: IPR011050; Pectin_lyas_1like.
DR Pfam: PF00544; Pec_lyase_C_1.
DR PRINTS: PR00807; AMBALALRGEN.
DR SMART: SM00656; Amb_all; 1.
KM
SQ SEQUENCE 418 AA; 45676 MW; 6FCEA055B929A260 CRC64;

Query Match          97.2%; Score 35; DB 2; Length 418;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGFCV 6
DB 171 AGGFCI 176

RESULT 22
AAQ84042 PRELIMINARY; PRT; 418 AA.
AC AAQ84042;
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DE Pectate lyase.
OS Malus domestica (Apple) (Malus sylvestris).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotsids I; Rosales; Rosaceae; Maloideae; Malus.
OX NCB1_TaxID=3750;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Royal Gala; TISSUE=Ripe fruit;
RA Goulae L.F., Oliveira C.M.;
RT "Isolation and characterization of a pectate lyase (PL) cDNA from ripening apples."
RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY376878; AAC84042.1; -.
KM
SQ SEQUENCE 418 AA; 45676 MW; 6FCEA055B929A260 CRC64;

Query Match          97.2%; Score 35; DB 2; Length 418;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGFCV 6
DB 171 AGGFCI 176

RESULT 23
027917 PRELIMINARY; PRT; 501 AA.
AC 027917;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Conserved protein.
OS OrderedLocustName=MTN1895;
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.

```

```

OX NCB1_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=96037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pochier B., Oiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
delta: functional analysis and comparative genomics."
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL: AE000941; AAB86355.1; -.
DR PIR: H69119; H69119.
DR InterPro: IPR006638; R1P3/MAB/NiFB.
DR InterPro: IPR007197; Radical SAM.
DR Pfam: PF04055; Radical SAM; 1.
DR SMART: SM00729; E1p3; 1.
KM
SQ SEQUENCE 501 AA; 56864 MW; E33198380BF6656C CRC64;

Query Match          97.2%; Score 35; DB 2; Length 501;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGFCV 6
DB 94 AGGFCI 99

RESULT 24
ACK1 HUMAN STANDARD; PRT; 1036 AA.
ID ACK1 HUMAN
AC Q07912; Q8N6U7; Q96H59;
DT 29-MAR-2004 (Ref. 43, Created)
DT 29-MAR-2004 (Ref. 43, Last sequence update)
DT 01-OCT-2004 (Ref. 45, Last annotation update)
DE Activated CDC42 kinase 1 (EC 2.7.1.112) (ACK-1).
GN Name=ACK1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH CDC42.
RC TISSUE=Hippocampus;
RX MEDLINE=93268389; PubMed=8497321;
RA Manser E., Leung T., Salihuddin H., Tan L., Lim L.;
RT "A non-receptor tyrosine kinase that inhibits the GTPase activity of p21cdc42."
RL Nature 363:364-367(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain, and uterus;
RX MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,
RA Stepleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshitsuki S., Carninci P., Prange C.,
RA Rana S.S., Loguigliano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

```





```

RESULT 26
Q8LH84 ID Q8LH84 PRELIMINARY; PRT; 94 AA.
AC Q8LH84;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein P030H06.127 (Hypothetical protein
DE P0475E07.107).
GN Name=P030H06.127; Synonyms=P0475E07.107;
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=3947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004395; BAC10207.1; -
DR EMBL; AP004568; BAC16099.1; -
KW Gramene; Q8LH84; -
KW Hypothetical protein.
SQ SEQUENCE 94 AA; 10446 MW; 86E47D0F6283429D CRC64;

Query Match 91.7%; Score 33; DB 2; Length 94;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCPV 6
Db 69 AGPCPL 74

RESULT 27
UL49 ID UL49 EBV STANDARD; PRT; 591 AA.
AC P14347;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Protein BFRF2.
GN Name=BFRF2;
OS Epstein-Barr virus (strain B95-8) (HHV-4) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=84270667; PubMed=6087149;
RX Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson J.J., Hatfull G., Hudson G.S., Satchwell S.C., Segun C.,
RA Tufnell P.S., Barrett B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
RN [2]
RP SIMILARITY: Belongs to the herpesviruses UL49 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V01555; CA24880.1; ALT_INIT.
CC InterPro; IPR004339; UL49.
CC Pfam; PF03117; Herpes_UL49_1; 1.

```

```

KW Early protein.
SQ SEQUENCE 591 AA; 63977 MW; BE63FF1E0721912E CRC64;

Query Match 91.7%; Score 33; DB 1; Length 591;
Best Local Similarity 83.3%; Pred. No. 7.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCPV 6
Db 333 AGPCPL 338

RESULT 28
Q777G6 ID Q777G6 PRELIMINARY; PRT; 591 AA.
AC Q777G6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE BFRF2 protein.
GN Name=BFRF2;
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=B95-8;
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R.J., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson J.J., Hatfull G.F., Hudson G.S., Satchwell S.C., Segun C.,
RA Tufnell P.S., Barrett B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=B95-8;
RX MEDLINE=88283646; PubMed=2840285;
RA Laux G., Perricaudet M., Farrell P.J.;
RT "A spliced Epstein-Barr virus gene expressed in immortalized
RT lymphocytes is created by circularization of the linear viral
RT genome.";
RL EMBO J. 7:769-774(1988).
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=B95-8;
RX MEDLINE=82014887; PubMed=6269068;
RA Arrand J.R., Rymo L., Walsh J.E., Bjorck E., Lindahl T., Griffin B.E.;
RT "Molecular cloning of the complete Epstein-Barr virus genome as a set
RT of overlapping restriction endonuclease fragments.";
RL Nucleic Acids Res. 9:299-3014(1981).
RN [4]
RP SEQUENCE FROM N.A.
RA STRAIN=B95-8;
RX MEDLINE=82059504; PubMed=7301588;
RA Korak M.;
RT "Possible role of flanking nucleotides in recognition of the AUG
RT initiator codon by eukaryotic ribosomes.";
RL Nucleic Acids Res. 9:5233-5252(1981).
RN [5]
RP SEQUENCE FROM N.A.
RA STRAIN=B95-8;
RX MEDLINE=83109311; PubMed=6296170;
RA Deininger P.L., Bankier A., Farrell P., Baer R., Barrett B.;
RT "Sequence analysis and in vitro transcription of portions of the
RT Epstein-Barr virus genome.";
RL J Cell. Biochem. 19:267-274(1982).
RN [6]
RP SEQUENCE FROM N.A.
RA STRAIN=B95-8;
RX MEDLINE=83169725; PubMed=6300857;
RA Farrell P.J., Deininger P.L., Bankier A., Barrett B.;
RT "Homologous upstream sequences near Epstein-Barr virus promoters.";

```

Proc. Natl. Acad. Sci. U.S.A. 80:1565-1569(1983).

[17] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RX MEDLINE=85035713; PubMed=6092825;

RA Bankier A.T., Deininger P.L., Farrell P.J., Barrell B.G.;

RT "Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8

RT Epstein-Barr virus.";

RL Mol. Biol. Med. 1:21-45(1983).

[18] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RX MEDLINE=85060424;

RA Segun C., Farrell P.J., Barrell B.G.;

RT "DNA sequence and transcription of the BamHI fragment B region of B95-

RT 8 Epstein-Barr virus.";

RL Mol. Biol. Med. 1:369-392(1983).

[19] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RX MEDLINE=83294686; PubMed=6310141;

RA Jeang K.T., Hayward S.D.;

RT "Organization of the Epstein-Barr virus DNA molecule. III. Location of

RT the P3HR-1 deletion junction and characterization of the NotI repeat

RT units that form part of the template for an abundant 12-O-

RT tetradecanoylphorbol-13-acetate-induced mRNA transcript.";

RL J. Virol. 48:135-148(1983).

[10] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RX MEDLINE=85060428; PubMed=6094955;

RA Bankier A.T., Deininger P.L., Satchwell S.C., Baer R., Farrell P.J.,

RA Barrell B.G.;

RT "DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-

RT Barr virus containing the terminal repeat sequences.";

RL Mol. Biol. Med. 1:425-445(1983).

[11] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RX MEDLINE=20331131; PubMed=10872327;

RA Farrell P.J., Bankier A., Segun C., Deininger P., Barrell B.G.;

RT "Latent and lytic cycle promoters of Epstein-Barr virus.";

RL EMBO J. 2:1331-1338(1983).

[12] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RX MEDLINE=84207939; PubMed=6327290;

RA Jones M.D., Foster L., Sheedy T., Griffin B.E.;

RT "The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion

RT similar to that observed in a non-transforming strain (P3HR-1) of the

RT virus.";

RL EMBO J. 3:813-821(1984).

[13] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RX MEDLINE=84236104; PubMed=6203743;

RA Biggin M., Farrell P.J., Barrell B.G.;

RT "Transcription and DNA sequence of the BamHI L fragment of B95-8

RT Epstein-Barr virus.";

RL EMBO J. 3:1083-1090(1984).

[14] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RX MEDLINE=84222045; PubMed=6328526;

RA Yates J., Warren N., Reisman D., Sugden B.;

RT "A cis-acting element from the Epstein-Barr viral genome that permits

RT stable replication of recombinant plasmids in latently infected

RT cells.";

RL Proc. Natl. Acad. Sci. U.S.A. 81:3806-3810(1984).

[15] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RX MEDLINE=84247360; PubMed=6330697;

RA Gibson T., Stockwell P., Ginsburg M., Barrell B.;

RT "Homology between two EBV early genes and HSV ribonucleotide reductase

RT and 38 genes.";

RL Nucleic Acids Res. 12:5087-5099(1984).

[16] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RX MEDLINE=87289053; PubMed=3039467;

RA Bodessot M., Pericaudet M.;

RT "Clustering alternative splice sites in Epstein-Barr virus RNAs.";

RL Nucleic Acids Res. 15:5887-5887(1987).

[17] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RX MEDLINE=91021036; PubMed=2171209;

RA Parker B.D., Bankier A., Satchwell S., Barrell B., Farrell P.J.;

RT "Sequence and transcription of Raji Epstein-Barr virus DNA spanning

RT the B95-8 deletion region.";

RL Virology 179:339-346(1990).

[18] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RA Hatfull G.F., Barrell B.G., Quinn J., McGeoch D.;

RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.

[19] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RA Blum U.K., Amon W., Farrell P.J.;

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ507799; CAD53400.1; -.

DR InterPro; IPR004339; UL49.

DR Pfam; PFO3117; Herpes\_UL49\_1; 1.

SO SEQUENCE 591 AA; 63977 MW; EE63FF1E0721912E CRC64;

Query Match 91.7%; Score 33; DB 2; Length 591;

Best Local Similarity 83.3%; Pred. NO. 7.6e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6

Db 333 AGPCV 338

RESULT 29

ID CAD53400 PRELIMINARY; PRT; 591 AA.

AC CAD53400.

DT 10-MAY-2004 (TREMURel. 27, Created)

DT 10-MAY-2004 (TREMURel. 27, Last sequence update)

DE BFRF2 protein.

DE BFRF2 protein.

GN Human herpesvirus 4 (Epstein-Barr virus).

OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Gammaherpesvirinae; Lymphocryptovirus.

OX NCBI\_TaxID=10376;

[1] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RX MEDLINE=82014887; PubMed=6269068;

RA Arrand J.R., Rymo U., Walsh J.E., Bjorck E., Lindahl T., Griffin B.E.;

RT "Molecular cloning of the complete Epstein-Barr virus genome as a set

RT of overlapping restriction endonuclease fragments.";

RL Nucleic Acids Res. 9:2999-3014(1981).

[2] SEQUENCE FROM N.A.

RC STRAIN=B95-8;

RX MEDLINE=82059504; PubMed=7301588;

RA Kozak M.;

RT "Possible role of flanking nucleotides in recognition of the AUG

RT initiator codon by eukaryotic ribosomes.";

RL Nucleic Acids Res. 9:5233-5252(1981).

[3]

RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=83109311; PubMed=6296170;  
 RA Deininger P.L., Bankier A., Farrell P., Baer R., Barrell B.;  
 RT "Sequence analysis and in vitro transcription of portions of the  
 RL Epstein-Barr virus genome.";  
 RN J. Cell. Biochem. 19:267-274 (1982).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=83169725; PubMed=6300857;  
 RA Farrell P.J., Deininger P.L., Bankier A., Barrell B.;  
 RT "Homologous upstream sequences near Epstein-Barr virus promoters.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:1565-1569 (1983).  
 [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=85035713; PubMed=6092825;  
 RA Bankier A.T., Deininger P.L., Farrell P.J., Barrell B.G.;  
 RT "Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8  
 RL Epstein-Barr virus.";  
 RN Mol. Biol. Med. 1:21-45 (1983).  
 [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=85060424; PubMed=6094953;  
 RA Seguin C., Farrell P.J., Barrell B.G.;  
 RT "DNA sequence and transcription of the BamHI fragment B region of B95-  
 RL 8 Epstein-Barr virus.";  
 RN Mol. Biol. Med. 1:369-392 (1983).  
 [7]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=83294686; PubMed=6310141;  
 RA Jeang K.T., Hayward S.D.;  
 RT "Organization of the Epstein-Barr virus DNA molecule. III. Location of  
 RL the P3HR-1 deletion junction and characterization of the NotI repeat  
 units that form part of the template for an abundant 12-O-  
 RL tetradecanoylphorbol-13-acetate-induced mRNA transcript.";  
 RN J. Virol. 48:1135-1148 (1983).  
 [8]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=85060428; PubMed=6094955;  
 RA Bankier A.T., Deininger P.L., Satchwell S.C., Baer R., Farrell P.J.,  
 RL Barrell B.G.;  
 RT "DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-  
 RL Barr virus containing the terminal repeat sequences.";  
 RN Mol. Biol. Med. 1:425-445 (1983).  
 [9]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=20331131; PubMed=10872327;  
 RA Farrell P.J., Bankier A., Seguin C., Deininger P., Barrell B.G.;  
 RL "Latent and lytic cycle promoters of Epstein-Barr virus.";  
 RN EMBO J. 2:1331-1338 (1983).  
 [10]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=84207939; PubMed=6327290;  
 RA Jones M.D., Foster L., Sheedy T., Griffin B.E.;  
 RT "The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion  
 RL similar to that observed in a non-transforming strain (P3HR-1) of the  
 RL virus.";  
 RN EMBO J. 3:813-821 (1984).  
 [11]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=84236104; PubMed=6203743;  
 RA Biggin M., Farrell P.J., Barrell B.G.;  
 RT "Transcription and DNA sequence of the BamHI L fragment of B95-8  
 RL Epstein-Barr virus.";  
 RN EMBO J. 3:1083-1090 (1984).

RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=84222045; PubMed=6328526;  
 RA Yates J., Warren N., Reisman D., Sugden B.;  
 RT "A cis-acting element from the Epstein-Barr viral genome that permits  
 RL stable replication of recombinant plasmids in latently infected  
 RL cells.";  
 RN Proc. Natl. Acad. Sci. U.S.A. 81:3806-3810 (1984).  
 [13]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=84247360; PubMed=6330697;  
 RA Gibson T., Stockwell P., Ginsburg M., Barrell B.;  
 RT "Homology between two EBV early genes and HSV ribonucleotide reductase  
 RL and 38k genes.";  
 RN Nucleic Acids Res. 12:5087-5099 (1984).  
 [14]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=84270657; PubMed=6087149;  
 RA Baer R.J., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
 RL Gibson T.J., Hatfull G.F., Hudson G.S., Satchwell S.C., Seguin C.,  
 RL Tuffnell P.S., Barrell B.G.;  
 RL "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";  
 RN Nature 310:207-211 (1984).  
 [15]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=87289053; PubMed=3039467;  
 RA Bodescot M., Perricaudet M.;  
 RT "Clustered alternative splice sites in Epstein-Barr virus RNAs.";  
 RL Nucleic Acids Res. 15:5887-5887 (1987).  
 [16]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=88283466; PubMed=2840285;  
 RA Laux G., Perricaudet M., Farrell P.J.;  
 RT "A spliced Epstein-Barr virus gene expressed in immortalized  
 RL lymphocytes is created by circularization of the linear viral  
 RL genome.";  
 RN EMBO J. 7:769-774 (1988).  
 [17]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=91021036; PubMed=2171209;  
 RA Parker B.D., Bankier A., Satchwell S., Barrell B., Farrell P.J.;  
 RL "Sequence and transcription of Raji Epstein-Barr virus DNA spanning  
 RL the B95-8 deletion region.";  
 RN Virology 179:339-346 (1990).  
 [18]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX Hatfull G.F., Barrell B.G., Quinn J., McGeoch D.;  
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
 [19]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX Blime U.K., Amon W., Farrell P.J.;  
 RT "Induction of Epstein-Barr virus late promoters on small plasmids in  
 RL the EBV late lytic cycle requires ori LYT.";  
 RN Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 [20]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX Farrell P.J.;  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ507799; CAD53400.1; -  
 SQ SEQUENCE 591 AA; 63977 MW; E63FF1E0721912B CRC64;  
 Query Match 91.7%; Score 33; DB 2; Length 591;  
 Best Local Similarity 83.3%; Pred. No. 7.6e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPCV 6  
 |||||  
 Db 333 AGPCV 338

## RESULT 30

06RKJ9 PRELIMINARY; PRT; 2434 AA.  
 ID O6RKJ9  
 AC O6RKJ9  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DE Polyketide synthase (Fragment).  
 GN Name=PKS1;  
 OS Botrytis cinerea (Noble rot fungus) (Botryotinia fuckeliana).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;  
 OC Helotiales; Sclerotiniaceae; Botryotinia.  
 OX NCBI\_TaxID=40559;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B05.10;  
 RX PubMed=14676319;  
 RA Krokken S., Glaes N.L., Taylor J.W., Yoder O.C., Turgeon B.G.;  
 RT "Phylogenomic analysis of type I polyketide synthase genes in  
 pathogenic and saprobic ascomycetes."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:15670-15675(2003).  
 DR EMBL; AY495606; AAR90237.1;  
 DR InterPro; IPR009081; ACP\_like.  
 DR InterPro; IPR001227; AC\_transferase.  
 DR InterPro; IPR002085; Adh\_zn\_family.  
 DR InterPro; IPR011032; GroES\_like.  
 DR InterPro; IPR000794; Ketoacyl\_synth.  
 DR InterPro; IPR006163; Pp\_bind.  
 DR InterPro; IPR000051; SAM\_bind.  
 DR Pfam; PF00698; ACP\_transfer\_1; 1.  
 DR Pfam; PF00107; ADH\_zinc\_N; 1.  
 DR Pfam; PF00109; ketoacyl-synt; 1.  
 DR Pfam; PF02801; ketoacyl-synt; 1.  
 DR PROSITE; PS50075; ACP\_DOMAIN; 1.  
 DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 1.  
 DR TRANSFERASE.  
 FT NON\_TER 1  
 FT NON\_TER 2434  
 SQ SEQUENCE 2434 AA; 269939 MW; 55887CFDC59CE245 CRC64;  
 Query Match 91.7%; Score 33; DB 2; Length 2434;  
 Best Local Similarity 83.3%; Pred. No. 2.7e+03;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPCV 6  
 |||||  
 Db 1511 SGPCV 1516

## RESULT 31

AAR90237 PRELIMINARY; PRT; 2434 AA.  
 ID AAR90237  
 AC AAR90237;  
 DT 02-MAR-2004 (TREMBlrel. 27, Created)  
 DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)  
 DE Polyketide synthase (Fragment).  
 GN PKS1  
 OS Botrytis cinerea (Noble rot fungus) (Botryotinia fuckeliana).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;  
 OC Helotiales; Sclerotiniaceae; Botryotinia.  
 OX NCBI\_TaxID=40559;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B05.10;  
 RX PubMed=14676319;  
 RA Krokken S., Glaes N.L., Taylor J.W., Yoder O.C., Turgeon B.G.;

RT "Phylogenomic analysis of type I polyketide synthase genes in  
 pathogenic and saprobic ascomycetes."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:15670-15675(2003).  
 DR EMBL; AY495606; AAR90237.1;  
 FT NON\_TER 1  
 FT NON\_TER 2434  
 SQ SEQUENCE 2434 AA; 269939 MW; 55887CFDC59CE245 CRC64;

Query Match 91.7%; Score 33; DB 2; Length 2434;  
 Best Local Similarity 83.3%; Pred. No. 2.7e+03;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPCV 6  
 |||||  
 Db 1511 SGPCV 1516

## RESULT 32

066009 PRELIMINARY; PRT; 73 AA.  
 ID 066009  
 AC 066009;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DE Phosphofructokinase I (EC 2.7.1.11) (Fragment).  
 GN Name=pfkI;  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 6057;  
 RX MEDLINE=20388102; PubMed=10932722;  
 RA Munoz M.E., Leborgne S., Bolivar F., Valle F.;  
 RT "Molecular cloning of the gene that codes for the pyruvate kinase of  
 Bacillus subtilis, primary characterization of a strain carrying this  
 RT gene insertionally inactivated."  
 RL Rev. Latinam. Microbiol. 39:129-140(1997).  
 DR EMBL; U73943; AAC12961.1;  
 DR HSSP; P00512; 3PFK.  
 DR GO; GO:0005945; C:6-phosphofructokinase complex; IEA.  
 DR GO; GO:0003872; F:6-phosphofructokinase activity; IEA.  
 DR GO; GO:0016301; F:kinase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006096; P:glycolysis; IEA.  
 DR InterPro; IPR000023; Pfrfructokinase.  
 DR Pfam; PF00365; PFK; 1.  
 DR KINASE; Transferase.  
 FT NON\_TER 1  
 FT NON\_TER 73  
 SQ SEQUENCE 73 AA; 7969 MW; 4768F4F9DB420AB8 CRC64;  
 Query Match 88.9%; Score 32; DB 2; Length 73;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
 |||||  
 Db 34 GGPCV 38

## RESULT 33

075WH6 PRELIMINARY; PRT; 77 AA.  
 ID 075WH6  
 AC 075WH6;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DE Peptide toxin 1.  
 OS Macrochele gigas (Spider).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
 OC Mysgalomorphae; Hexatheleidae; Macrochele.  
 OX NCBI\_TaxID=223896;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Venom gland;  
 RA Sarake H., Vallegas E., Corzo G.;  
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB121195; BAD13402.1; -  
 SQ SEQUENCE 77 AA; 8255 MW; B304C320CC9E1318 CRC64;

Query Match 88.9%; Score 32; DB 2; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GPCV 6  
 |||||  
 Db 50 GPCV 54

RESULT 34  
 BAD13402

ID BAD13402 PRELIMINARY; PRT; 77 AA.

AC BAD13402; (TREMBlrel. 27, Created)

DT 31-MAR-2004 (TREMBlrel. 27, Last sequence update)

DT 31-MAR-2004 (TREMBlrel. 27, Last annotation update)

DE Peptide toxin 1.

OS Macrochele gigas (Spider).

OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;

OC Mygalomorphae; Hexatheleidae; Macrochele.

OX NCBI\_TaxID=223896;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Venom gland;

RA Sarake H., Vallegas E., Corzo G.;

RL "cDNA encoding a peptide toxin of Macrochele gigas";

DR EMBL; AB121195; BAD13402.1; -

SQ SEQUENCE 77 AA; 8255 MW; B304C320CC9E1318 CRC64;

Query Match 88.9%; Score 32; DB 2; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GPCV 6  
 |||||  
 Db 50 GPCV 54

RESULT 35  
 Q7UYG4

ID Q7UYG4 PRELIMINARY; PRT; 91 AA.

AC Q7UYG4; (TREMBlrel. 25, Created)

DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Hypothetical protein.

GN OrderedLocusNames=RB629;

OS Rhodopirella ballica.

OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;

OC Planctomycetaceae; Pirellula.

OX NCBI\_TaxID=117;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=1; MEDLINE=22735913; PubMed=12835416;

RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,

RA Ludwig W., Gade D., Beck A., Borzym K., Helmann K., Rabus R.,

RA Schleuter H., Amann R., Reinhardt R.;

RL "Complete genome sequence of the marine planctomycete Pirellula sp.

RT strain 1";

RM Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).

DR EMBL; BX294134; CAD71678.1; -

KW Complete proteome; Hypothetical protein.

SQ SEQUENCE 91 AA; 9627 MW; 6FE56CA5D69BFB3B CRC64;

Query Match 88.9%; Score 32; DB 2; Length 91;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GPCV 6  
 |||||  
 Db 78 GPCV 82

RESULT 36  
 Q9CTQ0

ID Q9CTQ0 PRELIMINARY; PRT; 106 AA.

AC Q9CTQ0; (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE Mus musculus adult retina cDNA, RIKEN full-length enriched library,

DE clone:A930010C08 product:hypothetical protein, full insert sequence.

DE (Fragment).

GN Name=Ntng1;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Retina;

RX MEDLINE=99279253; PubMed=10349636;

RA Carninci P., Hayashizaki Y.;

RL "High-efficiency full-length cDNA cloning";

RM Meth. Enzymol. 303:19-44(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Retina;

RX MEDLINE=21085660; PubMed=11217851;

RA RIKEN FANTOM Consortium;

RL "Functional annotation of a full-length mouse cDNA collection";

RM Nature 409:685-690(2001).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Retina;

RX MEDLINE=20499374; PubMed=11042159;

RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,

RA Komno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;

RL "Normalization and substructure of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes";

RM Genome Res. 10:1617-1630(2000).

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Retina;

RX MEDLINE=20530913; PubMed=11076861;

RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,

RA Komno H., Akiyama J., Nishi K., Kitesuna T., Tashiro H., Itoh M.,

RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,

RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,

RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,

RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;

RL "RIKEN integrated sequence analysis (RISA) system-384-format

sequencing pipeline with 384 multicapillary sequencer";

RM Genome Res. 10:1757-1771(2000).

RN [6]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Retina;

RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,

```

RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y., to the EMBL/GenBank/DBJ databases.
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK020842; BAB32225.1; -.
DR MGD; MGI:1934028; Ntng1.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR GO; GO:0007409; P:axonogenesis; IDA.
KM Hypothetical protein.
FT NON_TER
SQ SEQUENCE 106 AA; 10899 MW; 575533A93108EBE7 CRC64;
Query Match 88.9%; Score 32; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGPC 5
Db 16 AGPC 20
RESULT 37
Q8BP6 PRELIMINARY; PRT; 108 AA.
ID Q8BP6;
AC Q8BP6;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched
DE library, clone:EI30203P04 product:hypothetical protein, full insert
DE sequence. (Fragment).
GN Name=Ntng1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID=10090;
[1]
RN RP STRAIN=C57BL/6J; TISSUE=EyeBall;
RC MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Mech. Enzymol. 303:19-44(1999).
[2]
RN RP STRAIN=C57BL/6J; TISSUE=EyeBall;
RC MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RN RP STRAIN=C57BL/6J; TISSUE=EyeBall;
RC The FANTOM Consortium;
RA "The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]
RN RP STRAIN=C57BL/6J; TISSUE=EyeBall;
RC MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";

```

```

RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EyeBall;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Kitaura T., Tashiro H., Itoh M.,
RA Sunit N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kasaiyagi K.,
RA Fujiwaka S., Inoue K., Togawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN RP STRAIN=C57BL/6J; TISSUE=EyeBall;
RC Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Kasukawa T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kaga H., Kato H., Kato S.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (Jul-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK053689; BAC35475.1; -.
DR MGD; MGI:1934028; Ntng1.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR GO; GO:0007409; P:axonogenesis; IDA.
KM Hypothetical protein.
FT NON_TER
SQ SEQUENCE 108 AA; 11167 MW; 186CA944B59AEP7A CRC64;
Query Match 88.9%; Score 32; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGPC 5
Db 18 AGPC 22
RESULT 38
Q7KUM2 PRELIMINARY; PRT; 119 AA.
ID Q7KUM2;
AC Q7KUM2;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE CG33259-PA.
GN ORFNames=CG33259;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCB1_TaxID=7227;
[1]
RN RP STRAIN=C57BL/6J; TISSUE=EyeBall;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Neilson C.R., Gaber G.L.,
RA Abrik J.F., Agbayani A., An H.U., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

```



RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Buttle K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heitman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195 (2000).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodson A.,  
 RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,  
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svitskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,  
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
 RT *melanogaster* euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 [3]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svitskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
 RT a genomic perspective.";  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 [4]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=22426069; PubMed=12537572;  
 RA Miara S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hardcastle P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
 [5]  
 RN SEQUENCE FROM N.A.  
 RP FLVBAS;  
 RG Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RG FLVBAS;  
 DR Submitted (MAR-2004) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AEO0531; AAS65002.1; -;  
 DR Interpro; IPR002919; Cyatich\_T1L.

DR Interpro; IPR006209; EGF\_like.  
 DR Pfam; PF01826; T1L; 1.  
 DR PROSITE; PS01186; EGF\_2; UNKNOWN 1.  
 SQ SEQUENCE 119 AA; 13124 MW; 7A6021C71297D287 CRC64;  
 Query Match 88.9%; Score 32; DB 2; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 2 GCPV 6  
 Db 57 GCPV 61  
 RESULT 39  
 AAS65002 PRELIMINARY; PRT; 119 AA.  
 AC AAS65002;  
 DT 01-APR-2004 (TREMBlrel. 27, Created)  
 DT 01-APR-2004 (TREMBlrel. 27, Last sequence update)  
 DT 01-APR-2004 (TREMBlrel. 27, Last annotation update)  
 DE CG33259-PA.  
 GN CG33259.  
 OS *Drosophila melanogaster* (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.M., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazey R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abri J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Baer A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bandaru D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Buttle K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195 (2000).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,

RA Patel S., Adams M., Champe M., Dugan S.P., Friese E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
 RA Paclab J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskaas R., Tabor P.E., Wen K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstock G., Wheeler S.E., Myers E.W., Gibbs R.A., Rubin G.M.,  
 RT "Finishing a whole-genome shotgun release 3 of the Drosophila  
 RT melanogaster euchromatic genome sequence."  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Miya S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu I., Berman B.P.,  
 RA Bertencourt B.R., Celinker S.E., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.,  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review."  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskaas R.,  
 RA Patel S., Friese E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase:  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 [6]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase:  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003531; AAS65002.1; - 7A6021C71297D2B7 CRC64;  
 SQ SEQUENCE 119 AA; 13124 MW; 7A6021C71297D2B7 CRC64;  
 Query Match 88.9%; Score 32; DB 2; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GPCV 6  
 DB 57 GPCV 61

RESULT 40  
 Q6SHL1 PRELIMINARY; PRT; 132 AA.  
 AC Q6SHL1;  
 DT 05-JUL-2004 (TREMBLrel. 27, Created)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
 DE Hypothetical protein  
 GN ORFNames=EBAC750-09G06.25;  
 OS uncultured bacterium 314.  
 OC Bacteria; environmental samples.  
 OX NCB1\_TaxID=257387;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA DeLong E.F.;  
 RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., DeLong E.F.;  
 RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY458634; AAR37610.1; -  
 DR InterPro; IPR006683; Thioestr\_supf.  
 DR Pfam; PF03061; 4HBT; 1.

KM Hypothetical protein.  
 SQ SEQUENCE 132 AA; 14323 MW; 9359C91A9247C4E5 CRC64;  
 Query Match 88.9%; Score 32; DB 2; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GPCV 6  
 DB 71 GPCV 75

Search completed: December 30, 2004, 15:13:58  
 Job time : 8.93462 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: December 30, 2004, 14:59:21 ; Search time 0.211484 Seconds  
(without alignments)  
2729.763 Million cell updates/sec

Title: US-10-719-385-26

Perfect score: 36

Sequence: 1 AGGPCV 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

1: PIR 79:.\*  
2: p1r1:.\*  
3: p1r3:.\*  
4: p1r4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	100.0	407	2	T34535
2	36	100.0	444	2	F64420
3	36	100.0	469	2	C69460
4	35	97.2	383	2	S49928
5	35	97.2	383	2	S67477
6	35	97.2	501	2	H69119
7	35	97.2	1091	2	H69119
8	33	91.7	657	2	S32991
9	32	88.9	208	2	T41555
10	32	88.9	254	2	T12338
11	32	88.9	280	2	C70617
12	32	88.9	292	2	S11870
13	32	88.9	322	2	T03866
14	32	88.9	374	2	JC2124
15	32	88.9	374	2	JC2124
16	32	88.9	386	2	A41950
17	32	88.9	398	2	C95278
18	32	88.9	404	2	T05556
19	32	88.9	404	2	S12209
20	32	88.9	426	2	S22586
21	32	88.9	426	2	F96994
22	32	88.9	437	2	JC4929
23	32	88.9	450	1	WABPT5
24	32	88.9	454	2	A97181
25	32	88.9	462	1	C71296
26	32	88.9	519	2	T43533
27	32	88.9	531	2	T40480
28	32	88.9	535	2	T40887
29	32	88.9	587	2	JC5669

30	32	88.9	687	2	T27421	hypothetical prote
31	32	88.9	699	2	A43674	US4 protein - huma
32	32	88.9	1220	2	A56136	tagged protein pre
33	32	88.9	1319	2	S55598	segment protein 0
34	32	88.9	1408	2	S16148	gene serrate prote
35	32	88.9	2135	2	T14602	variant-specific s
36	31	86.1	48	2	S29216	neurotoxin Tx2 - s
37	31	86.1	49	2	S29215	neurotoxin Tx2 - s
38	31	86.1	291	2	T20083	hypothetical prote
39	31	86.1	291	2	T26576	hypothetical prote
40	31	86.1	570	4	B44282	retrovirus-related
41	31	86.1	2437	2	S42612	transmembrane prot
42	31	86.1	2824	2	T22759	hypothetical prote
43	30	83.3	96	2	S32182	hypothetical prote
44	30	83.3	96	2	T08518	hypothetical prote
45	30	83.3	104	2	J01757	hypothetical 11.6K
46	30	83.3	213	2	S17815	bacteriophage phi
47	30	83.3	396	2	JC5022	UDP-galactose tran
48	29	80.6	44	2	S54144	tmal protein - mou
49	29	80.6	84	2	H96495	unknown protein, 4
50	29	80.6	101	2	A97623	hypothetical prote
51	29	80.6	105	2	A13464	hypothetical prote
52	29	80.6	110	2	G72540	hypothetical prote
53	29	80.6	111	2	I50622	Crot protein - chi
54	29	80.6	124	2	T48833	hypothetical prote
55	29	80.6	202	2	B45512	cold-regulated pro
56	29	80.6	208	2	T46927	hypothetical prote
57	29	80.6	209	2	C81691	conserved hypotnet
58	29	80.6	210	2	S55671	hypothetical prote
59	29	80.6	244	2	F75520	indoleoleglycerol-
60	29	80.6	255	2	C81301	hypothetical prote
61	29	80.6	273	2	T49602	related to homoti
62	29	80.6	346	2	I48185	gene alx3 protein
63	29	80.6	351	2	T31920	hypothetical prote
64	29	80.6	353	2	D69422	F420-nonreducing h
65	29	80.6	374	2	T22683	hypothetical prote
66	29	80.6	390	2	JC7162	Golgi UDP-galactos
67	29	80.6	432	2	A28913	regulatory protein
68	29	80.6	483	2	E49343	salicylaldehyde de
69	29	80.6	503	2	T48825	hypothetical prote
70	29	80.6	513	1	ACCHD1	nicotinic acetylch
71	29	80.6	554	2	C70512	hypothetical prote
72	29	80.6	662	2	D40228	neurexin II-beta p
73	29	80.6	673	4	F40201	artifac-warnin s
74	29	80.6	1715	2	C40228	neurexin II-alpha
75	29	80.6	2533	2	T28675	alpha-51D immobili
76	29	80.6	2533	2	T28674	alpha-51D immobili
77	29	80.6	3078	2	T28432	variant-specific s
78	29	80.6	4351	2	T00252	MEGF1 protein - ra
79	28	77.8	63	2	G87562	hypothetical prote
80	28	77.8	63	2	S25772	testis-specific pr
81	28	77.8	64	2	C87540	hypothetical prote
82	28	77.8	68	2	S25775	testis-specific pr
83	28	77.8	74	2	S25773	testis-specific pr
84	28	77.8	75	2	AE1708	hypothetical prote
85	28	77.8	80	1	AE1737	hypothetical prote
86	28	77.8	80	1	S01498	pancreatic secreto
87	28	77.8	82	2	D81056	hypothetical prote
88	28	77.8	88	2	T47139	hypothetical prote
89	28	77.8	89	2	S71555	proteinase inhibit
90	28	77.8	89	2	J02361	wheat aluminum ind
91	28	77.8	90	2	A87553	Peab protein (imp
92	28	77.8	103	2	C34770	ORF3 protein - sai
93	28	77.8	121	2	A10552	conserved hypotnet
94	28	77.8	127	2	E87434	hypothetical prote
95	28	77.8	130	2	S41571	blaeficidin-S deam
96	28	77.8	134	2	T49623	hypothetical prote
97	28	77.8	138	2	S74831	hypothetical prote
98	28	77.8	141	2	G72641	hypothetical prote
99	28	77.8	143	2	B72627	hypothetical prote
100	28	77.8	154	2	H97860	hypothetical prote
101	28	77.8	171	2	B70408	conserved hypotnet
102	28	77.8	172	1	RFMB	phycoerythrocyanin

103	28	77.8	173	2	T36170	hypothetical prote
104	28	77.8	179	2	S58124	hypothetical prote
105	28	77.8	186	2	A28401	agglutinin isolect
106	28	77.8	189	2	T02792	hypothetical prote
107	28	77.8	199	2	T03991	probable transcrip
108	28	77.8	201	2	A47096	response regulator
109	28	77.8	203	2	D8116	RNA methyltransfer
110	28	77.8	212	2	T05936	agglutinin isolect
111	28	77.8	212	2	S09623	agglutinin isolect
112	28	77.8	213	1	AEWT2	agglutinin isolect
113	28	77.8	214	1	S6476	cysteine proteinas
114	28	77.8	214	2	T50577	hypothetical prote
115	28	77.8	215	2	T01349	hypothetical prote
116	28	77.8	219	2	E70431	hypothetical prote
117	28	77.8	228	2	T01350	hypothetical prote
118	28	77.8	232	2	E87520	hypothetical prote
119	28	77.8	242	1	RHMT	thyroidalderin precu
120	28	77.8	245	1	E69010	conserved hypotet
121	28	77.8	246	1	VCFV37	coat protein, sp37
122	28	77.8	246	2	S23734	env polypotein ev
123	28	77.8	246	2	A60398	env polypotein -
124	28	77.8	256	2	T45970	hypothetical prote
125	28	77.8	257	2	A43362	env polypotein -
126	28	77.8	257	2	C84850	hypothetical prote
127	28	77.8	259	2	E84866	probable beta-expa
128	28	77.8	261	2	B81904	probable RNA methy
129	28	77.8	265	2	B48151	sperm tail protein
130	28	77.8	280	1	A39357	connexin-crossreac
131	28	77.8	283	2	P91285	hypothetical prote
132	28	77.8	294	2	B86177	hypothetical prote
133	28	77.8	296	2	H70585	hypothetical prote
134	28	77.8	298	2	T03990	probable transcrip
135	28	77.8	307	2	T38206	probable phosphopr
136	28	77.8	312	2	D37202	hypothetical prote
137	28	77.8	319	2	A83898	forminoglutamase
138	28	77.8	322	2	G83852	hypothetical prote
139	28	77.8	323	2	G84325	hypothetical prote
140	28	77.8	324	2	A69934	thioredoxin reduct
141	28	77.8	325	2	D87669	Sua5/Xc10/Yrdc fam
142	28	77.8	328	2	H89926	hypothetical prote
143	28	77.8	334	2	A48151	sperm tail protein
144	28	77.8	335	2	T44498	hypothetical prote
145	28	77.8	339	2	A47611	env polypotein -
146	28	77.8	340	2	E97822	hypothetical prote
147	28	77.8	342	2	F87562	transcription regu
148	28	77.8	343	2	T06273	benzochadiazole-i
149	28	77.8	345	2	F75454	peptide ABC transp
150	28	77.8	347	2	B47611	env polypotein -

## ALIGNMENTS

RESULT 1  
T34535  
hypothetical protein DKFZp434H0717.1 - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T34535  
R/Biocheck: H.; Biocheck, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, October 1999  
A/Reference number: Z21539  
A/Accession: T34535  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-407 <BLO>  
A/Cross-references: UNIPROT:Q9UPF4; EMBL:AL122102  
A/Experimental source: adult testis; clone DKFZp434H0717  
C/Genetics:  
A/Note: DKFZp434H0717.1

Query Match 100.0%; Score 36; DB 2; Length 407;  
Best Local Similarity 100.0%; Pred. No. 25;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGGPCV 6  
Db 280 AGGPCV 285

RESULT 2  
F64420  
hypothetical protein M0966 - Methanococcus jannaschii  
C/Species: Methanococcus jannaschii  
C/Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
C/Accession: F64420  
R/Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, A.;  
rgon, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hult, M.A.  
Science 273, 1058-1073, 1996  
A/Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Frazer, C.M.; Smith, H.O.; Woese, C.  
A/Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.  
A/Reference number: A64300; PMID:96337999; PMID:8688087  
A/Accession: F64420  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-444 <BUL>  
A/Cross-references: UNIPROT:Q58376; GB:U67539; GB:L77117; NID:G1591619; PIDN:AAB89868.1;  
C/Genetics:  
A/Map position: FOR899868-901202  
A/Start codon: GNG  
C/Superfamily: hypothetical protein M0966

Query Match 100.0%; Score 36; DB 2; Length 444;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6  
Db 88 AGGPCV 93

RESULT 3  
C69460  
conserved hypothetical protein AF1684 - Archaeoglobus fulgidus  
C/Species: Archaeoglobus fulgidus  
C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C/Accession: C69460  
R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,  
.. Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A/Authors: Usterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.A.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeor  
A/Reference number: A69250; PMID:98049343; PMID:9389475  
A/Accession: C69460  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-469 <KLE>  
A/Cross-references: UNIPROT:Q28589; GB:AE000987; GB:AE000782; NID:G2669310; PIDN:AAB8956;  
C/Superfamily: hypothetical protein M0966

Query Match 100.0%; Score 36; DB 2; Length 469;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6  
Db 104 AGGPCV 109

RESULT 4  
F420-nonreducing hydrogenase I vhog - Methanosarcina mazei  
C/Species: Methanosarcina mazei

C:/Date: 13-Jan-1995 #sequence\_revision 12-May-1995 #text\_change 09-Jul-2004  
C:/Accession: S49928; S67474  
R:/Depositor: U. Blaust, M.; Lentes, S.; Herzberg, C.; Gottechaik, G.  
A:/Description: Analysis of the vhcGAC and vhcGAC operons from Methanosarcina mazei strain  
A:/Reference number: S49928  
A:/Accession: S49928  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-383 <DEP>  
A:/Cross-references: UNIPROT:Q50248; EMBL:X82940; NID:G599897; PIDN:CAA58113.1; PID:G5998  
R:/Depositor: U. Blaust, M.; Lentes, S.; Herzberg, C.; Gottechaik, G.  
Eur. J. Biochem. 227, 261-269, 1995  
A:/Title: Analysis of the vhcGAC and vhcGAC operons from Methanosarcina mazei strain Goel  
A:/Reference number: S67474; MUID:95154297; PMID:7851393  
A:/Accession: S67474  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-383 <DEP>  
A:/Cross-references: EMBL:X82940; NID:G599897; PIDN:CAA58113.1; PID:G599898  
C:/Genetics:  
A:/Gene: vhcG  
C:/Superfamily: hydrogenase (NiFe) small chain

Query Match 97.2%; Score 35; DB 2; Length 383;  
Best Local Similarity 83.3%; Pred. No. 36;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 323 AGPCCI 328

RESULT 5  
S67477  
P420-nonreducing hydrogenase (EC 1.12.99.-) vhcG precursor - Methanosarcina mazei  
C:/Species: Methanosarcina mazei  
C:/Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 09-Jul-2004  
C:/Accession: S67477  
R:/Depositor: U. Blaust, M.; Lentes, S.; Herzberg, C.; Gottechaik, G.  
Eur. J. Biochem. 227, 261-269, 1995  
A:/Title: Analysis of the vhcGAC and vhcGAC operons from Methanosarcina mazei strain Goel  
A:/Reference number: S67474; MUID:95154297; PMID:7851393  
A:/Accession: S67477  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-383 <DEP>  
A:/Cross-references: UNIPROT:Q50225; EMBL:X83112; NID:G602581; PIDN:CAA58176.1; PID:G6025  
C:/Superfamily: hydrogenase (NiFe) small chain,  
C:/Keywords: oxidoreductase

Query Match 97.2%; Score 35; DB 2; Length 383;  
Best Local Similarity 83.3%; Pred. No. 36;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 323 AGPCCI 328

RESULT 6  
H69119  
hypothetical protein MTH1895 - Methanobacterium thermoautotrophicum (strain Delta H)  
C:/Species: Methanobacterium thermoautotrophicum  
C:/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C:/Accession: H69119  
R:/Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
Kl., S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A:/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funcn  
A:/Reference number: A69000; MUID:98037514; PMID:9371463  
A:/Accession: H69119

A:/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:/Molecule type: DNA  
A:/Residues: 1-501 <MTH>  
A:/Cross-references: UNIPROT:Q27917; GB:AE000941; GB:AE000666; NID:G2623025; PIDN:AA8635;  
A:/Experimental source: strain Delta H  
C:/Genetics:  
A:/Gene: MTH1895  
A:/Start codon: TTG  
C:/Superfamily: hypothetical protein M0966

Query Match 97.2%; Score 35; DB 2; Length 501;  
Best Local Similarity 83.3%; Pred. No. 46;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 94 AGPCCI 99

RESULT 7  
S33596  
protein-tyrosine kinase (EC 2.7.1.112) - human  
C:/Species: Homo sapiens (man)  
C:/Date: 02-Dec-1993 #sequence\_revision 01-Sep-1995 #text\_change 16-Aug-2004  
C:/Accession: S33596  
R:/Manser, E.; Leung, T.; Salihuddin, H.; Tan, L.; Lim, L.  
Nature 363, 364-367, 1993  
A:/Title: A non-receptor tyrosine kinase that inhibits the GTPase activity of p21 (cdc42).  
A:/Reference number: S33596; MUID:93268389; PMID:8497321  
A:/Accession: S33596  
A:/Status: preliminary  
A:/Molecule type: mRNA  
A:/Residues: 1-1091 <MAN>  
A:/Cross-references: UNIPROT:Q07912; EMBL:L13738; NID:G307304; PIDN:AAA53570.1; PID:G30730  
C:/Superfamily: protein kinase homology  
C:/Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase  
F180-448/Domain: protein kinase homology <KIN>  
F188-196/Region: protein kinase ATP-binding motif

Query Match 97.2%; Score 35; DB 2; Length 1091;  
Best Local Similarity 83.3%; Pred. No. 89;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 892 AGPCCI 897

RESULT 8  
S32991  
hypothetical protein - human herpesvirus 4  
C:/Species: human herpesvirus 4, Epstein-Barr virus  
C:/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Jan-1995  
C:/Accession: S32991  
R:/Farrell, P.J.  
submitted to the EMBL Data Library, March 1988  
A:/Reference number: S32973  
A:/Accession: S32991  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-657 <PAR>  
A:/Cross-references: EMBL:V01555

Query Match 91.7%; Score 33; DB 2; Length 657;  
Best Local Similarity 83.3%; Pred. No. 1,46+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 399 AGPCCL 404

RESULT 9

T41555  
 hypothetical protein SPCC70.09c - fission yeast (Schizosaccharomyces pombe)  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C/Accession: T41555  
 R/Wood, V.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Harris, D.  
 submitted to the EMBL Data Library, June 1998  
 A/Reference number: Z22001  
 A/Accession: T41555  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-208 <MO>  
 A/Cross-references: UNIPROT:O94243; EMBL:AL023794; PIDD:CAA19359.1; GSPDB:GN00068; SPDB:  
 A/Experimental source: strain 972h-; contmid c70  
 A/Genetics:  
 A/Gene: SPDB:SPCC70.09c  
 A/Map position: 3  
 A/Introns: 48/1

Query Match 88.9%; Score 32; DB 2; Length 208;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
 |||||  
 DB 5 AGGPC 9

## RESULT 10

T12338  
 L-ascorbate peroxidase (EC 1.11.1.11) - common ice plant  
 C/Species: Mesembryanthemum crystallinum (common ice plant)  
 C/Date: 23-Jul-1998 #sequence\_revision 23-Jul-1999 #text\_change 12-Jul-2004  
 C/Accession: T12338  
 R/Michalowski, C.B.; Bohner, H.J.  
 submitted to the EMBL Data Library, July 1998  
 A/Description: A putative ascorbate peroxidase from the common ice plant Mesembryanthemum  
 A/Reference number: Z17496  
 A/Accession: T12338  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-254 <MIC>  
 A/Cross-references: UNIPROT:O81604; EMBL:AF079513; NID:G3377754; PID:G3377755  
 C/Superfamily: peroxidase  
 C/Keywords: heme; iron; metalloprotein; oxidoreductase  
 F/160/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 88.9%; Score 32; DB 2; Length 254;  
 Best Local Similarity 66.7%; Pred. No. 91;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPCV 6  
 |||||  
 DB 107 SGGPCI 112

## RESULT 11

C70617  
 probable transcription regulator RV0144 - Mycobacterium tuberculosis (strain H37Rv)  
 C/Species: Mycobacterium tuberculosis  
 C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
 C/Accession: C70617  
 R/Cole, S.T.; Broach, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A/Reference number: A70500; MUID:98295987; PMID:9634230  
 A/Accession: C70617  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-280 <COL>

A/Cross-references: UNIPROT:P96821; GB:Z92770; GB:AL123456; NID:G3261720; PIDD:CA807050.1  
 A/Experimental source: strain H37Rv  
 A/Genetics:  
 A/Gene: RV0144

Query Match 88.9%; Score 32; DB 2; Length 280;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
 |||||  
 DB 33 AGGPC 37

## RESULT 12

S11870  
 peroxidase (EC 1.11.1.7) - cucumber (fragment)  
 C/Species: Cucumis sativus (cucumber)  
 C/Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
 C/Accession: S11870  
 R/Morgans, P.H.; Callahan, A.M.; Dunn, L.J.; Abeles, F.B.  
 Plant Mol. Biol. 14, 715-725, 1990  
 A/Title: Isolation and sequencing of cDNA clones encoding ethylene-induced putative perox  
 A/Reference number: S11870; MUID:91346662; PMID:2102850  
 A/Accession: S11870  
 A/Molecule type: mRNA  
 A/Residues: 1-292 <MOR>  
 A/Cross-references: UNIPROT:P19135; GB:M32742; NID:G167516; PIDD:AAA3121.1; PID:G167517  
 A/Note: 231-Leu was also found  
 C/Superfamily: peroxidase  
 C/Keywords: chromoprotein; glycoprotein; heme; iron; metalloprotein; oxidoreductase  
 F/7-86/Diulfide bonds: #status predicted  
 F/34/Active site: Arg #status predicted  
 F/38/164/Binding site: heme iron (His) (axial ligands) #status predicted  
 F/40-45/Diulfide bonds: #status predicted  
 F/66/119/179/Binding site: carbonylate (Asn) (covalent) #status predicted  
 F/92-288/Diulfide bonds: #status predicted  
 F/171-199/Diulfide bonds: #status predicted

Query Match 88.9%; Score 32; DB 2; Length 292;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
 |||||  
 DB 107 AGGPC 111

## RESULT 13

T03686  
 peroxidase (EC 1.11.1.7) - common tobacco  
 C/Species: Nicotiana tabacum (common tobacco)  
 C/Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004  
 C/Accession: T03686  
 R/Osakabe, K.; Kawai, S.; Katayama, Y.; Morohoshi, N.  
 submitted to the EMBL Data Library, June 1992  
 A/Description: Nucleotide sequence for the genomic DNA encoding the anionic peroxidase ge  
 A/Reference number: Z15008  
 A/Accession: T03686  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-322 <OSA>  
 A/Cross-references: UNIPROT:Q42964; EMBL:DL1396; PIDD:BA01992.1  
 A/Experimental source: strain NK326  
 A/Genetics:  
 A/Introns: 66/3; 128/3; 184/1  
 C/Superfamily: peroxidase  
 C/Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase  
 F/33-109/Diulfide bonds: #status predicted  
 F/60/Active site: Arg #status predicted  
 F/64/187/Binding site: heme iron (His) (axial ligands) #status predicted  
 F/66-69/Diulfide bonds: #status predicted  
 F/115-318/Diulfide bonds: #status predicted

F:194-226/D1enuflide bonds: #status predicted

Query March 88.9%; Score 32; DB 2; Length 322;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
|||||  
Db 130 AGGPC 134

#### RESULT 14

JC2123

major allergen Cry j I precursor (clone PCCI-2-2) - Japanese cedar  
C:Species: Cryptomeria japonica (Japanese cedar)  
C>Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 09-Jul-2004  
C:Accession: JC2123; PC2065  
R:Song, T.; Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.  
Biochem. Biophys. Res. Commun. 199, 619-625, 1994  
A>Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japanese  
A:Reference number: JC2123; MUID:94183234; PMID:8135802  
A:Accession: JC2123  
A:Molecule type: mRNA  
A:Cross-references: UNIPROT:P18632; GB:D26544; NID:G493631; PID:BA05542.1; PID:G493632  
A:Experimental source: pollen

A:Accession: PC2065  
A:Molecule type: protein  
A:Residues: 22-53;58-81;219-232;236-258;299-307;346-372 <SO2>  
A>Note: the authors described carbohydrate binding site for residue 279  
C:Superfamily: pectate lyase LAT59  
C:Keywords: glycoprotein; pollen  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-374/Product: major allergen Cry j I (clone PCCI-2-2) #status predicted <MAT>  
F:158,191,293,354/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query March 88.9%; Score 32; DB 2; Length 374;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
|||||  
Db 125 GGPCV 129

#### RESULT 15

JC2124

major allergen Cry j I precursor (clone PCCI-15) - Japanese cedar  
C:Species: Cryptomeria japonica (Japanese cedar)  
C>Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 09-Jul-2004  
C:Accession: JC2124  
R:Song, T.; Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.  
Biochem. Biophys. Res. Commun. 199, 619-625, 1994  
A>Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japanese  
A:Reference number: JC2123; MUID:94183234; PMID:8135802  
A:Accession: JC2124  
A:Molecule type: mRNA  
A:Residues: 1-374 <SON>  
A:Cross-references: UNIPROT:P18632; GB:D26544; NID:G493633; PID:BA05543.1; PID:G493634  
A:Experimental source: pollen  
A>Note: the authors described carbohydrate binding site for residue 279  
C:Superfamily: pectate lyase LAT59  
C:Keywords: glycoprotein; pollen  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-374/Product: major allergen Cry j I (clone PCCI-15) #status predicted <MAT>  
F:158,191,293,354/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query March 88.9%; Score 32; DB 2; Length 374;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
|||||

Db 125 GGPCV 129

#### RESULT 16

A41950

retrovirus-related hypothetical protein 1 - Trypanosoma cruzi retrotransposon  
C:Species: Trypanosoma cruzi  
C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Jan-1999  
C:Accession: A41950  
R:Villanueva, M.S.; Williams, S.P.; Beard, C.B.; Richards, F.F.; Aksoy, S.  
Mol. Cell. Biol. 11, 6139-6148, 1991  
A>Title: A new member of a family of site-specific retrotransposons is present in the sp1  
A:Reference number: A41950; MUID:92049344; PMID:1179380  
A:Accession: A41950  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-386 <VIL>  
A:Cross-references: GB:M62862; NID:G162247; PID:G162248  
A>Note: sequence extracted from NCBI backbone (NCBIN:66378, NCBI:66379)

Query March 88.9%; Score 32; DB 2; Length 386;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
|||||  
Db 346 AGGPC 350

#### RESULT 17

C95278

hypothetical protein Sma0247 [imported] - Sinorhizobium meliloti (strain 1021) magaplaemmu  
C:Species: Sinorhizobium meliloti  
C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: C95278  
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowser,  
J.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Welle, D.H.; Yeh, K.C.;  
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
A>Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti  
A:Reference number: A95262; MUID:21396509; PMID:11481432  
A:Accession: C95278  
A:Molecule type: DNA  
A:Status: preliminary  
A:Residues: 1-398 <KIR>  
A:Cross-references: UNIPROT:Q930R4; GB:AE006469; PID:AAK64789.1; PID:G14523198; GSPDB:GT  
A:Experimental source: strain 1021, megaplaemid pSYMA  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
F.; Chain, P.; Cowie, A.; Davies, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
hehault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.C.  
A>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: Sma0247  
A:Genome: plasmid

Query March 88.9%; Score 32; DB 2; Length 398;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
|||||  
Db 37 AGGPC 41

#### RESULT 18

T05556

pectate lyase (EC 4.2.2.2) F22K18.20 - Arabidopsis thaliana  
N:Alternate names: protein F22K18.20  
C:Species: Arabidopsis thaliana (mouse-ear cress)



C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999  
C/Accession: T05556  
R/Beyan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hohnsbeil, J.; Mewes, H.W.; Mayer, K.F.  
submitted to the Protein Sequence Database, February 1999  
A/Reference number: 215419  
A/Accession: T05556  
A/Molecule type: DNA  
A/Residues: 1-404 <BEV>  
A/Cross-references: EMBL:AL035356  
A/Experimental source: cultivar Columbia; BAC clone F22K18  
C/Genetics:  
A/Map position: 4  
A/Introns: 37/2; 287/2  
A/Note: F22K18.20  
C/Superfamily: pectate lyase LATS9  
C/Keywords: carbon-oxygen lyase

Query Match 88.9%; Score 32; DB 2; Length 404;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 159 SGGPCI 164

## RESULT 19

S12209  
pectate lyase (EC 4.2.2.2) - tomato  
C/Species: Lycopersicon esculentum (tomato)  
C/Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004  
C/Accession: S12209  
R/Budeller, K.A.; Smith, A.G.; Gasser, C.S.  
Mol. Gen. Genet. 224, 183-192, 1990  
A/Title: Regulation of a stylar transmitting tissue-specific gene in wild-type and trans  
A/Reference number: S12209; MUID:91117185; PMID:2277637  
A/Accession: S12209  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-404 <BD>  
A/Cross-references: UNIPROT:P24396; GB:X55193; NID:919161; PIDN:CAA38979.1; PID:919162  
C/Superfamily: pectate lyase LATS9  
C/Keywords: carbon-oxygen lyase

Query Match 88.9%; Score 32; DB 2; Length 404;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 155 SGGPCI 160

## RESULT 20

S22586  
homeotic protein EVX1 - human  
C/Species: Homo sapiens (man)  
C/Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 16-Aug-2004  
C/Accession: S22586  
R/Falletta, A.; d'Esposito, M.; Rambaldi, M.; Acampora, D.; Balsaniore, S.; Stornaiuolo,  
Nucleic Acids Res. 19, 6541-6545, 1991  
A/Title: Isolation and mapping of EVX1, a human homeobox gene homologous to even-skipped  
A/Reference number: S22586; MUID:92093615; PMID:1684419  
A/Accession: S22586  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-407 <FAL>  
A/Cross-references: UNIPROT:P49640; EMBL:X60655; NID:931278; PIDN:CAA43062.1; PID:977357  
C/Superfamily: homeobox homology  
C/Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F184-240/Domain: homeobox homology <HGX>

Query Match 88.9%; Score 32; DB 2; Length 407;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 5  
Db 343 AGGPC 347

## RESULT 21

P96994  
uncharacterized conserved protein CAC0769 [imported] - Clostridium acetobutylicum  
C/Species: Clostridium acetobutylicum  
C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C/Accession: P96994  
R/Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J. Bacteriol. 183, 4823-4838, 2001  
A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A/Reference number: A96900; MUID:21359325; PMID:21359325  
A/Accession: P96994  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-426 <KUR>  
A/Cross-references: UNIPROT:Q97KX7; GB:AE001437; PIDN:AAK78745.1; PID:915023654; GSPDB:G  
A/Experimental source: Clostridium acetobutylicum ATCC824  
C/Genetics:  
A/Genes: CAC0769

Query Match 88.9%; Score 32; DB 2; Length 426;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 151 SGGPCI 156

## RESULT 22

JC4929  
transcription factor E2F1 - human  
N/Alternate names: Ap12 protein; retinoblastoma-associated protein; retinoblastoma-bindin  
C/Species: Homo sapiens (man)  
C/Date: 22-Oct-1996 #sequence\_revision 01-Nov-1996 #text\_change 09-Jul-2004  
C/Accession: JC4929; A45032; A42998; A42997; I54091  
R/Neuman, E.; Sellers, W.R.; McNeill, J.A.; Lawrence, J.B.; Kaelin Jr., W.G.  
Gene 173, 163-169, 1996  
A/Title: Structure and partial genomic sequence of the human E2F1 gene.  
A/Reference number: JC4929; MUID:97082961; PMID:8964493  
A/Accession: JC4929  
A/Status: preliminary; nucleic acid sequence not shown  
A/Molecule type: DNA  
A/Residues: 1-437 <NEU>  
A/Cross-references: UNIPROT:Q01094; GB:U47675; NID:91594281; GB:U47676; NID:91594282; GB  
A/Experimental source: Placenta  
R/Shan, B.; Zhu, X.; Chen, P.L.; Durfee, T.; Yang, Y.; Sharp, D.; Lee, W.H.  
Mol. Cell. Biol. 12, 5620-5631, 1992  
A/Title: Molecular cloning of cellular genes encoding retinoblastoma-associated proteins  
A/Reference number: A45032; MUID:93078763; PMID:1448092  
A/Accession: A45032

A/Molecule type: mRNA  
A/Residues: 1-CRQRPPAERDRALAAECRHGPAPPPPTTAAANVS, 1-312, 'S', 314-321, 'N', 323-328, 'T',  
A/Cross-references: GB:S49592; NID:9260573; PIDN:AA824289.1; PID:9260574  
A/Note: sequence extracted from NCBI backbone (NCBI:P119095)  
A/Note: the authors are uncertain whether Met-1 is the initiator or whether translation  
R/Helin, K.; Lees, J.A.; Vidal, M.; Dyson, N.; Harlow, E.; Pattae, A.  
Cell 70, 337-350, 1992  
A/Title: A cDNA encoding a PRB-binding protein with properties of the transcription fact  
A/Reference number: A42998; MUID:92346720; PMID:1638634  
A/Accession: A42998  
A/Molecule type: mRNA  
A/Residues: 1-312, 'S', 314-321, 'N', 323-328, 'T', 330-437 <HEL>  
A/Cross-references: GB:M96577; NID:9181917; PIDN:AA35782.1; PID:9181918  
A/Experimental source: Nalm 6 pre-B leukemia cell line

A/Note: sequence extracted from NCBI backbone (NCBIN:110015, NCBI:110016)  
R:Kellin Jr., W.G.; Krek, W.; Sellers, W.R.; Decaprio, J.A.; Ajchenbaum, F.; Fuchs, C.S.  
Cell 70, 351-364, 1992  
A/Title: Expression cloning of a cDNA encoding a retinoblastoma-binding protein with E2F  
A/Reference number: A42997; MUID:92346721; PMID:1638635  
A/Accession: A42997  
A/Molecule type: mRNA  
A/Residues: 1-312, 'S', 314-321, 'N', 323-328, 'T', 330-437 <KAB>  
A/Cross-references: GB:U13851; NID:9585713  
A/Experimental source: Akata cells; expression vector pGEX-2TK  
A/Note: sequence extracted from NCBI backbone (NCBIN:110018, NCBI:110019)  
R:Johnson, D.G.; Ohnishi, K.; Nevins, J.R.  
Genes Dev. 8, 1514-1525, 1994  
A/Title: Autoregulatory control of E2F1 expression in response to positive and negative  
A/Reference number: 154091; MUID:95047311; PMID:7958836  
A/Accession: 154091  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-88, 'R', 'T', 122-123, 'TPR', 127, 'ORR', 297-299, 'PRR', 308-309, 'RA', 312, 'C' <RES>  
A/Cross-references: GB:674230; NID:9712816; PIDN:AD14150.1; PID:94261850  
C/Genetics:  
A/Status: GDB:E2F1  
A/Cross-references: GDB:134661; OMIM:189971  
A/Map position: 20q11-20q11  
A/Introns: 87/3 118/1 191/2 242/2 280/3 356/1  
C/Keywords: DNA binding; transcription factor  
F:67-108/Region: cyclin box #status predicted  
F:118-190/Domain: DNA binding #status predicted <DNA>  
F:191-241/Region: 7-residue repeats

Query Match 88.9%; Score 32; DB 2; Length 437;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPC 5  
Db 8 AGPC 12

RESULT 23  
WABPT5  
gene D10 protein - phage T5  
C/Species: phage T5  
A/Note: host Escherichia coli  
C/Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C/Accession: S01931  
R:Kallman, A.V.; Kryukov, V.M.; Bayev, A.A.  
Nucleic Acids Res. 16, 10353-10354, 1988  
A/Title: The nucleotide sequence of the region of bacteriophage T5 early genes D10-D15.  
A/Reference number: S01931; MUID:89057468; PMID:3057441  
A/Accession: S01931  
A/Status: translation not shown  
A/Molecule type: DNA  
A/Residues: 1-450 <KAL>  
A/Cross-references: UNIPROT:P11107; EMBL:X12930; NID:915407; PIDN:CAA31398.1; PID:915405  
C/Genetics:  
A/Status: D10  
C/Superfamily: phage T5 gene D10 protein  
C/Keywords: DNA binding; early protein

Query Match 88.9%; Score 32; DB 1; Length 450;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GPCV 6  
Db 347 GPCV 351

RESULT 24  
A97181  
heme biosynthesis (nirx-2) family protein [imported] - Clostridium acetobutylicum  
C/Species: Clostridium acetobutylicum

C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C/Accession: A97181  
R:Nolling, J.; Berton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A/Reference number: A96900; MUID:21359325; PMID:21359325  
A/Accession: A97181  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-454 <KUR>  
A/Cross-references: UNIPROT:Q97GTE; GB:AE001437; PIDN:AKR80236.1; PID:915025284; GSPDB:G  
A/Experimental source: Clostridium acetobutylicum ATCC824  
C/Genetics:  
A/Status: CAC2279

Query Match 88.9%; Score 32; DB 2; Length 454;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GPCV 6  
Db 335 GPCV 339

RESULT 25  
C71296  
glycine-tRNA ligase (EC 6.1.1.14) (glys) - syphilis spirochete  
C/Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: C71296  
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McDor  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A/Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A/Reference number: A71250; MUID:98332770; PMID:9665876  
A/Accession: C71296  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-462 <COL>  
A/Cross-references: UNIPROT:O83678; GB:AE001241; GB:AE000520; NID:93322969; PIDN:AAC2657C  
A/Experimental source: strain Nichols  
C/Genetics:  
A/Status: TP0672  
C/Superfamily: Mycoplasma genitalium glycine-tRNA ligase  
C/Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 88.9%; Score 32; DB 1; Length 462;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPC 5  
Db 103 AGPC 107

RESULT 26  
T43533  
hexose transport protein Gnt2 - fission yeast (Schizosaccharomyces pombe)  
C/Species: Schizosaccharomyces pombe  
C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C/Accession: T43533  
R:Heiland, S.; Hofer, M.; Radovanovic, N.; Lichtenberg-Frate, H.  
submitted to the EMBL Data Library, August 1997  
A/Description: Multiple hexose transporters of Schizosaccharomyces pombe.  
A/Reference number: Z22549  
A/Accession: T43533  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-519 <HEI>  
A/Cross-references: UNIPROT:O74969; EMBL:AF017180; PIDN:AAB70519.1  
A/Experimental source: strain 972h(-)

C/Genetics:  
A/Gene: ght2  
C/Superfamily: maltose transport protein MAL61

Query Match 88.9%; Score 32; DB 2; Length 519;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
|||||  
Db 240 AGGPC 244

## RESULT 27

T40480  
hexose transporter - fission yeast (Schizosaccharomyces pombe)  
C/Species: Schizosaccharomyces pombe  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T40480  
R/Beck, A.; Reinhardt, R.; Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, May 1997  
A/Reference number: Z21932  
A/Accession: T40480  
A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA  
A/Residues: 1-531 <BEC>  
A/Cross-references: UNIPROT:O74969; EMBL:AL023706; PIDN:CAA19288.1; GSPDB:GN00067; SPDB:  
C/Genetics:  
A/Gene: ght2; SPDB:SPBC484.08  
A/Map position: 2  
C/Superfamily: maltose transport protein MAL61

Query Match 88.9%; Score 32; DB 2; Length 531;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
|||||  
Db 252 AGGPC 256

## RESULT 28

T40887  
hexose transporter ght6 [similarity] - fission yeast (Schizosaccharomyces pombe)  
C/Species: Schizosaccharomyces pombe  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T40887; T50484  
R/Mood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.  
submitted to the EMBL Data Library, September 1998  
A/Reference number: Z21954  
A/Accession: T40887  
A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA  
A/Residues: 1-535 <MOO>  
A/Cross-references: UNIPROT:O74849; EMBL:AL031164; PIDN:CAA21117.1; GSPDB:GN00068; SPDB:  
A/Experimental source: strain 972h(-); cosmid c1235  
R/Halland, S.; Hoefler, M.; Radovanovic, N.; Winderickx, J.; Lichtenberg, H.  
submitted to the EMBL Data Library, October 1998  
A/Description: Multiple hexose transporters of Schizosaccharomyces pombe.  
A/Reference number: Z25086  
A/Accession: T50484  
A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA  
A/Residues: 1-535 <HEI>  
A/Cross-references: EMBL:AF098076; PIDN:AA64976.1  
A/Experimental source: strain 972h(-)  
C/Genetics:  
A/Gene: SPDB:SPCC1235.13; ght6  
A/Map position: 3  
C/Superfamily: maltose transport protein MAL61

Query Match 88.9%; Score 32; DB 2; Length 535;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
|||||  
Db 250 AGGPC 254

## RESULT 29

JC5669  
Ca2+/calmodulin-dependent protein kinase kinase (EC 2.7.1.1-) beta chain - rat  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 20-Nov-1997 #sequence\_revision 20-Nov-1997 #text\_change 16-Aug-2004  
C/Accession: JC5669; PC4493  
R/Kitani, T.; Okuno, S.; Fujisawa, H.  
J. Biochem. 122, 243-250, 1997  
A/Title: Molecular cloning of Ca2+/calmodulin-dependent protein kinase kinase beta.  
A/Reference number: JC5669; MIMD:97420710; PMID:9276695  
A/Accession: JC5669

A/Molecule type: mRNA  
A/Residues: 1-587 <KIT>  
A/Cross-references: UNIPROT:O88831; GB:AB018081; NID:G3702720; PIDN:BAA33524.1; PID:dl03;  
A/Experimental source: cerebellar  
A/Accession: PC4493  
A/Molecule type: protein  
A/Residues: 425-501 <KIT>  
A/Experimental source: cerebellar  
C/Superfamily: protein kinase homology  
C/Keywords: ATP; phosphotransferase  
F/162-445/Domain: protein kinase homology <KIN>  
F/170-194/Domain: ATP-binding #status predicted <ATP>  
F/480-493/Domain: calmodulin-binding #status predicted <CAB>

Query Match 88.9%; Score 32; DB 2; Length 587;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGPCV 6  
|||||  
Db 557 GGPCV 561

## RESULT 30

T27421  
hypothetical protein Y76A2B.3 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T27421  
R/Steward, C.  
submitted to the EMBL Data Library, October 1998  
A/Reference number: Z20363  
A/Accession: T27421

A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-667 <WIL>  
A/Cross-references: UNIPROT:Q9XMD1; EMBL:AL032658; PIDN:CAA21744.1; GSPDB:GN00021; CESP:  
A/Experimental source: clone Y76A2B  
C/Genetics:  
A/Gene: CESP:Y76A2B.3  
A/Map position: 3  
A/Intons: 46/3; 84/3; 130/1; 220/3; 363/3; 494/3; 582/2  
C/Superfamily: human long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homology  
F/126-669/Domain: acetate-CoA ligase homology <ACL>

Query Match 88.9%; Score 32; DB 2; Length 687;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGPCV 6  
|||||  
Db 88 GGPCV 92

## RESULT 31

C43674

USA protein - human herpesvirus 2 (strain HG52)

C:Species: human herpesvirus 2

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004

C:Accession: C43674

R:McGeoch, D.J.; Moss, H.W.M.; McNab, D.; Frame, M.C.

J. Gen. Virol. 68, 19-38, 1987

A:Title: DNA sequence and genetic content of the HindIII 1 region in the short unique co

nventional comparisons.

A:Reference number: A43674; MUID:87111457; PMID:3027242

A:Accession: C43674

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-699 &lt;MG&gt;

A:Cross-references: UNIPROT:P13290; EMBL:X04798; NID:G53900; PID:CAA28490.1; PID:G53904

C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

## Query Match

Best Local Similarity 100.0%; Score 32; DB 2; Length 699;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GPCV 6

Db 215 GPCV 219

## RESULT 32

A56136

Jagged protein precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 28-Apr-1995 #sequence\_revision 28-Apr-1995 #text\_change 21-Jul-2003

C:Accession: A56136

R:Lindeberg, C.E.; Shawber, C.J.; Boulter, J.; Weinmaster, G.

Cell 80, 909-917, 1995

A:Title: Jagged: a mammalian ligand that activates Notch1.

A:Reference number: A56136; MUID:95211842; PMID:7697721

A:Accession: A56136

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1220 &lt;LIN&gt;

A:Cross-references: GB:L38483

F:379-410/Domain: EGF homology &lt;EGF1&gt;

F:492-523/Domain: EGF homology &lt;EGF&gt;

F:634-665/Domain: EGF homology &lt;EGF2&gt;

## Query Match

Best Local Similarity 100.0%; Score 32; DB 2; Length 1220;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPC 5

Db 89 AGPC 93

## RESULT 33

S55598

Legument protein 03 - equine herpesvirus 2

C:Species: equine herpesvirus 2

C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 09-Jul-2004

C:Accession: S55598

R:Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.

J. Mol. Biol. 249, 520-528, 1995

A:Title: The DNA sequence of equine herpesvirus 2.

A:Reference number: S55594; MUID:95302501; PMID:7783207

A:Accession: S55598

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1319 &lt;TEL&gt;

A:Cross-references: UNIPROT:Q66609; GB:U20824; NID:G695172; PID:AA13791.1; PID:G695176

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

Best Local Similarity 100.0%; Pred. No. 3.8e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPC 5

Db 116 AGPC 120

## RESULT 34

S16148

Gene serrate protein precursor - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 31-Dec-1991 #sequence\_revision 02-Aug-1994 #text\_change 09-Jul-2004

C:Accession: S16148; S16878; A36666

R:Thomas, U.; Speicher, S.A.; Knust, E.

Development 11, 749-761, 1991

A:Title: The Drosophila gene Serrate encodes an EGF-like transmembrane protein with a co

A:Reference number: S16148; MUID:91347903; PMID:1840519

A:Accession: S16148

A:Molecule type: mRNA

A:Residues: 1-1408 &lt;THO1&gt;

A:Cross-references: UNIPROT:P18168; EMBL:X56811

R:Thomas, U.

submitted to the EMBL Data Library, November 1990

A:Reference number: S16878

A:Accession: S16878

A:Molecule type: mRNA

A:Residues: 1-1351, T, 1353-1408 &lt;THO2&gt;

A:Cross-references: EMBL:X56811; NID:G8563; PID:G8564

R:Fleming, R.J.; Scottgale, T.N.; Diederich, R.J.; Artavanis-Tsakonas, S.

Genes Dev. 4, 2188-2201, 1990

A:Title: The gene Serrate encodes a putative EGF-like transmembrane protein essential fo

A:Reference number: A36666; MUID:91099666; PMID:2125287

A:Accession: A36666

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-15, 20-26, A, 28-1408 &lt;FLE&gt;

A:Cross-references: GB:M5759; NID:G158605; PID:G158606

C:Genetics:

A:Gene: FlyBase:Ser

A:Cross-references: FlyBase:FBgn0004197

C:Keywords: glycoprotein; transmembrane protein

F:1-84/Domain: signal sequence #status predicted &lt;SIG&gt;

F:85-1408/Product: gene serrate protein #status predicted &lt;MAT&gt;

F:85-1221/Domain: extracellular #status predicted &lt;EXT&gt;

F:1283-316/Domain: EGF homology &lt;EG01&gt;

F:319-348/Domain: EGF homology &lt;EG02&gt;

F:355-388/Domain: EGF homology &lt;EG03&gt;

F:395-488/Domain: EGF homology #status atypical &lt;EG04&gt;

F:495-526/Domain: EGF homology &lt;EG05&gt;

F:533-608/Domain: EGF homology #status atypical &lt;EG06&gt;

F:615-645/Domain: EGF homology &lt;EG07&gt;

F:652-683/Domain: EGF homology &lt;EG08&gt;

F:690-720/Domain: EGF homology &lt;EG09&gt;

F:727-796/Domain: EGF homology #status atypical &lt;EG10&gt;

F:803-834/Domain: EGF homology &lt;EG11&gt;

F:841-876/Domain: EGF homology &lt;EG12&gt;

F:883-914/Domain: EGF homology &lt;EG13&gt;

F:921-952/Domain: EGF homology &lt;EG14&gt;

F:997-1060/Region: cysteine-rich

F:1122-1246/Domain: transmembrane #status predicted &lt;TM1&gt;

F:1247-1408/Domain: intracellular #status predicted &lt;INT&gt;

F:1152,196,247,331,412,452,558,739,965,977,1004,1030,1050/Binding site: carbohydrate (Asn

Query Match

Best Local Similarity 100.0%; Score 32; DB 2; Length 1408;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPC 5

Db 496 AGPC 500

RESULT 35  
T14602  
variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragments)  
C/Species: Plasmodium falciparum  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: T14602  
R/Voss, T.S.; Felger, I.; Weis, N.; Beck, H.P.  
submitted to the EMBL Data Library, February 1998  
A/Description: Identification of a conserved 5' flanking region of Plasmodium falciparum  
A/Reference number: Z18158  
A/Accession: T14602  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-2135 <VOS>  
A/Cross-references: UNIPROT:O61077; EMBL:AF050740; NID:G2944094; PID:G2944095; PIDN:AACG  
C/Genetics:  
A/Gene: varph17

Query Match 88.9%; Score 32; DB 2; Length 2135;  
Best Local Similarity 100.0%; Pred. No. 5.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPC 5  
Db 897 AGPC 901

RESULT 36  
S29216  
neurotoxin Tx2 - spider (Phonotria nigriverter)  
C/Species: Phonotria nigriverter  
C/Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004  
C/Accession: S29216  
R/do Nascimento Cordeiro, M.; Ribeiro Diniz, C.; do Carmo Valentim, A.; von Bickstedt, V  
FEBS Lett. 310, 153-156, 1992  
A/Title: The purification and amino acid sequences of four Tx2 neurotoxins from the venom  
A/Reference number: S29214; MUID:93011905; PMID:1397265  
A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 1-48 <COR>  
A/Cross-references: UNIPROT:P29425  
C/Superfamily: curratotoxin

Query Match 86.1%; Score 31; DB 2; Length 48;  
Best Local Similarity 80.0%; Pred. No. 33;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GPCV 6  
Db 26 GPCV 30

RESULT 37  
S29215  
neurotoxin Tx2 - spider (Phonotria nigriverter)  
C/Species: Phonotria nigriverter  
C/Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004  
C/Accession: S29215; B39305  
R/do Nascimento Cordeiro, M.; Ribeiro Diniz, C.; do Carmo Valentim, A.; von Bickstedt, V  
FEBS Lett. 310, 153-156, 1992  
A/Title: The purification and amino acid sequences of four Tx2 neurotoxins from the venom  
A/Reference number: S29214; MUID:93011905; PMID:1397265  
A/Accession: S29215  
A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 1-49 <COR>  
A/Cross-references: UNIPROT:P29424  
R/Rezende Jr., L.; Cordeiro, M.N.; Oliveira, E.B.; Diniz, C.R.  
Toxicon 29, 1225-1233, 1991  
A/Title: Isolation of neurotoxic peptides from the venom of the 'armed' spider Phonotria  
A/Reference number: A59305; MUID:9219603; PMID:1801316  
A/Accession: B39305

A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 1-11 <REZ>  
C/Superfamily: curratotoxin  
C/Keywords: neurotoxin; venom

Query Match 86.1%; Score 31; DB 2; Length 49;  
Best Local Similarity 80.0%; Pred. No. 34;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GPCV 6  
Db 26 GPCV 30

RESULT 38  
T20083  
hypothetical protein C50B6.4 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T20083  
R/percy, C.  
submitted to the EMBL Data Library, October 1996  
A/Reference number: Z19222  
A/Accession: T20083  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-291 <WIL>  
A/Cross-references: UNIPROT:O17666; EMBL:Z81050; PIDN:CAB02849.1; GSPDB:GN00023; CESP:CS  
C/Genetics:  
A/Gene: CESP:C50B6.4  
A/Experimental source: clone C50B6  
A/Map position: 5  
A/Introns: 47/3

Query Match 86.1%; Score 31; DB 2; Length 291;  
Best Local Similarity 80.0%; Pred. No. 1.6e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GPCV 6  
Db 127 GPCV 131

RESULT 39  
T26576  
hypothetical protein Y2H9A.3 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T26576  
R/Wallis, J.  
submitted to the EMBL Data Library, December 1998  
A/Reference number: Z20237  
A/Accession: T26576  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-291 <WIL>  
A/Cross-references: UNIPROT:Q9U2P5; EMBL:AL021448; PIDN:CAA16275.1; GSPDB:GN00023; CESP:  
A/Experimental source: clone Y2H9A  
C/Genetics:  
A/Gene: CESP:Y2H9A.3  
A/Map position: 5  
A/Introns: 47/3

Query Match 86.1%; Score 31; DB 2; Length 291;  
Best Local Similarity 80.0%; Pred. No. 1.6e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GPCV 6  
Db 127 GPCV 131

## RESULT 40

B44282

retrovirus-related env polyprotein pseudogene - human

N/Alternate names: coat polyprotein

N/Contains: coat protein gp70; coat protein p20E

C/Species: Homo sapiens (man)

C/Date: 17-Feb-1994 #sequence\_revision 04-Jan-1996 #text\_change 09-Jul-2004

C/Accession: B44282

R/Hirose, Y.; Takamatsu, M.; Harada, F.

Virology 192, 52-61, 1993

A/Title: Presence of env genes in members of the RVL-H family of human endogenous retro

A/Reference number: A44282; MUID:93297138; PMID:8517031

A/Accession: B44282

A/Status: conceptual translation of pseudogene

A/Molecule type: mRNA

A/Residues: 1-570 &lt;HIR&gt;

A/Cross-references: UNIPROT:Q9N2K0; GB:D10083

C/Genetics:

A/Gene: env

C/Keywords: coat protein; polyprotein; pseudogene

F:1-372/Domain: coat protein gp70 &lt;CGP&gt;

F:370/Region: opal stop codon

F:373-569/Domain: coat protein p20E &lt;CPP&gt;

F:436-448/Region: immunosuppressive peptide

## Query Match

86.1%; Score 31; DB 4; Length 570;

Best Local Similarity 80.0%; Pred. No. 2.8e+02;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GPCV 6

DB 454 GPCV 458

Search completed: December 30, 2004, 15:14:59  
Job time : 5.21148 secs

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2004, 14:59:21 : Search time 61.7885 Seconds  
(without alignments)  
2729.763 Million cell updates/sec

Title: US-10-719-385-2

Perfect score: 9007  
Sequence: 1 MIRSRTSVLSFCRSSREL.....PESQEPILQVQAVRHMQR 1753

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database :  
1: p1r1.\*  
2: p1r2.\*  
3: p1r3.\*  
4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	234	2.6	1628	2 T13682	nucleoporin - fis
2	215.5	2.4	1655	2 T06075	hypothetical prote
3	214	2.4	1655	2 S47446	nucleoporin Np188
4	192.5	2.1	1647	2 T41267	hypothetical prote
5	192.5	2.1	1683	2 S56811	probable membrane
6	184.5	2.0	2670	2 T37919	GCM1 homolog - fis
7	175.5	1.9	1073	2 T06074	hypothetical prote
8	175	1.9	2110	2 H96803	unknown protein T5
9	171.5	1.9	2104	2 T38774	myosin-3 heavy cha
10	169	1.9	2698	2 B96671	similar to transla
11	167	1.9	4385	2 T29042	hypothetical prote
12	163.5	1.8	2076	2 T28915	hypothetical prote
13	155	1.7	2491	2 A57036	talin - slime mold
14	154.5	1.7	3227	2 T37964	probable ubiquitin
15	151	1.7	3225	2 T00637	hypothetical prote
16	151	1.7	3225	2 I52300	giantin - human
17	151	1.7	3259	1 A56539	giantin - human
18	151	1.7	4464	2 D87755	protein T21E2.4 (
19	149.5	1.7	1334	2 T19493	hypothetical prote
20	147	1.6	2954	2 T14156	kinesin-related pr
21	146.5	1.6	1868	2 S48938	hypothetical prote
22	145.5	1.6	2848	2 T32550	hypothetical prote
23	144.5	1.6	2594	2 A35774	kinase-related pro
24	144	1.6	2693	2 A40743	lpi receptor, xrp
25	143	1.6	2549	2 S45340	FKBP-rapamycin-as
26	142.5	1.6	2700	2 D88450	protein F21H1.2 (
27	142	1.6	2549	2 A54837	rapamycin/FKBP12 c
28	141.5	1.6	2114	2 E96505	hypothetical prote
29	141	1.6	4924	2 T50176	probable peptide s

30	140.5	1.6	1826	2 H86502	excinuclease ABC s
31	140	1.6	2777	2 D96746	hypothetical prote
32	139.5	1.5	4096	2 A57099	DNA-activated prot
33	138.5	1.5	4672	2 A48126	translation activa
34	138.5	1.5	3744	2 S46715	hypothetical prote
35	138	1.5	2241	2 S09811	hypothetical prote
36	137	1.5	2764	2 T39499	neurofibromin - fr
37	137	1.5	2802	2 T13947	neurofibromin - fr
38	135.5	1.5	2335	2 T40186	probable phosphati
39	135	1.5	1570	2 T42735	TBP-interacting pr
40	135	1.5	1230	2 AC2012	hypothetical prote
41	134.5	1.5	3788	2 T13960	beige protein homo
42	134	1.5	2048	2 C84609	hypothetical prote
43	134	1.5	2470	2 S57085	1-phosphatidylinos
44	133	1.5	1825	2 T42725	actin binding prot
45	133	1.5	1885	2 T30847	actin binding prot
46	133	1.5	2033	2 T30849	actin binding prot
47	133	1.5	2297	2 AB2494	hypothetical prote
48	132.5	1.5	1067	2 T39449	probable importin
49	132.5	1.5	1837	2 T41023	probable nuclear p
50	132	1.5	851	2 T38173	probable phosphati
51	132	1.5	1649	2 T39938	hypothetical prote
52	131	1.5	1896	2 T01490	hypothetical prote
53	131	1.5	4128	2 JC6306	protein kinase (BC
54	130.5	1.4	1920	2 A53188	pericenturin - me
55	130	1.4	839	1 S71553	endopeptidase C1p
56	130	1.4	2802	2 T13945	neurofibromin - fr
57	130	1.4	4056	2 H96599	protein F14J16.10
58	129.5	1.4	1796	2 AC1895	serine/threonine k
59	129.5	1.4	1888	2 T39009	hypothetical prote
60	129	1.4	821	2 T40994	hypothetical prote
61	129	1.4	1108	2 T06633	hypothetical prote
62	129	1.4	1696	2 T32617	hypothetical prote
63	129	1.4	1966	2 T32552	hypothetical prote
64	129	1.4	2327	2 T31733	hypothetical prote
65	129	1.4	2761	2 T29285	hypothetical prote
66	128.5	1.4	1826	2 D72130	exinuclease ABC
67	128.5	1.4	3092	2 S46009	GTPase-activating
68	128.5	1.4	3796	2 T18514	lysosomal trafficking
69	128	1.4	1232	2 T21018	hypothetical prote
70	128	1.4	1321	2 T42842	P-glycoprotein sis
71	128	1.4	1321	2 T42842	bile salt transpor
72	128	1.4	2276	2 T00076	hypothetical prote
73	127.5	1.4	1018	2 T40253	hypothetical prote
74	127.5	1.4	2149	2 C96695	ribulose biphosph
75	127.5	1.4	2513	2 G96536	hypothetical prote
76	127.5	1.4	3433	1 S28381	utrophin - human
77	127	1.4	549	2 D82483	sensor histidine k
78	127	1.4	2535	2 T04824	hypothetical prote
79	126.5	1.4	1630	2 T40217	hypothetical prote
80	126.5	1.4	2330	1 KR1MW	genome polyprotein
81	126.5	1.4	2671	2 A49873	inositol 1,4,5-tri
82	126.5	1.4	2783	2 T31431	inositol 1,4,5-tri
83	126.5	1.4	4845	2 T31067	BIR repeat contain
84	126	1.4	971	2 T10678	hypothetical prote
85	126	1.4	2701	2 S17996	inositol-triphosph
86	126	1.4	3944	2 T19997	hypothetical prote
87	125.5	1.4	1064	2 A41542	adenylate cyclase
88	125.5	1.4	1254	2 T47141	hypothetical prote
89	125.5	1.4	2092	2 S30026	genome polyprotein
90	125.5	1.4	2149	2 S18676	genome polyprotein
91	125	1.4	1325	2 T42722	male-enhanced anti
92	125	1.4	4151	2 T13734	groovin gene prote
93	124.5	1.4	1660	2 A84647	hypothetical prote
94	124.5	1.4	3588	2 T40485	suflactin syntheta
95	124	1.4	821	1 S76330	endopeptidase C1p
96	124	1.4	958	2 T10679	hypothetical prote
97	124	1.4	1089	2 S53978	PEB1 protein yea
98	124	1.4	1196	2 S65245	translation elonga
99	124	1.4	3449	2 T01083	hypothetical prote
100	123.5	1.4	1824	2 T07589	disease resistance
101	123.5	1.4	2359	2 T03094	A-kinase anchor pr
102	123	1.4	1101	2 G83637	hypothetical prote

103	123	1.4	1274	2	T25024	hypothetical prote
104	123	1.4	1299	2	AH2090	two-component hybr
105	123	1.4	2331	2	S44054	genome polypeptid
106	123	1.4	2749	1	ACHSIT	inositol 1,4,5-tri
107	122.5	1.4	1823	2	S28974	vitellinogenin precu
108	122.5	1.4	1946	2	AC2141	serine/threonine k
109	122.5	1.4	4391	2	A38096	plecan precursor
110	122.5	1.4	4466	1	S17231	dynein beta heavy
111	122	1.4	710	2	AH1109	anaerobic ribonuc
112	122	1.4	768	2	HE5013	yea protein - Ebc
113	122	1.4	1180	2	A11939	two-component hybr
114	122	1.4	1607	2	T03022	map kinase kinase
115	122	1.4	1806	2	T23298	probable helicase
116	122	1.4	1953	2	T0642	phosphotidylinosit
117	122	1.4	2812	2	T43271	hypothetical prote
118	122	1.4	3738	2	T05501	hypothetical prote
119	122	1.4	6486	2	T31076	hypothetical prote
120	121.5	1.3	930	2	T00403	hypothetical prote
121	121.5	1.3	1018	2	T18249	conserved hypothet
122	121.5	1.3	1744	2	F86161	F1003.10 protein -
123	121	1.3	729	2	A81038	hypothetical prote
124	121	1.3	742	2	D85882	hypothetical prote
125	121	1.3	790	2	T19040	hypothetical prote
126	121	1.3	1577	2	T15851	hypothetical prote
127	120.5	1.3	1043	2	G72073	hypothetical prote
128	120.5	1.3	1383	2	G67273	hypothetical prote
129	120.5	1.3	2182	2	T14330	calcineurin inhibi
130	120.5	1.3	2493	2	T40540	hypothetical prote
131	120.5	1.3	2617	2	A82136	peptide synthetase
132	120.5	1.3	3795	2	T00831	hypothetical prote
133	120.5	1.3	4910	2	S64942	probable membrane
134	120	1.3	639	2	F82382	methy-accepting c
135	120	1.3	1420	2	S54471	probable membrane
136	120	1.3	1755	2	T51532	hypothetical prote
137	120	1.3	2337	2	T40577	probable phosphati
138	120	1.3	2354	2	T13288	mel-41 protein - f
139	119.5	1.3	1033	2	F81595	hypothetical prote
140	119.5	1.3	1043	2	G86550	hypothetical prote
141	119.5	1.3	1207	2	T19041	rae GTPase-activat
142	119.5	1.3	1228	2	S59681	probable membrane
143	119.5	1.3	1422	2	B71437	probable resistance
144	119.5	1.3	1493	2	H71445	hypothetical prote
145	119.5	1.3	1574	1	A38454	myosin MYO2 - yea
146	119.5	1.3	1769	2	S53378	probable membrane
147	119	1.3	714	2	A82458	hypothetical prote
148	119	1.3	2713	2	A55713	histocol 1,4,5-tri
149	119	1.3	3788	2	T10851	lysosomal traffick
150	118.5	1.3	956	2	T19046	ras GTPase-activat

## ALIGNMENTS

RESULT 1  
 T43682  
 nucleoporin - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
 C:Accession: T43682  
 R:Whalen, W.A.; Yoon, J.H.; Shen, R.; Dhar, R.  
 A:Title: Regulation of mRNA export by nutritional status in fission yeast.  
 A:Reference number: 222629; PMID:99318821; PMID:10388805  
 A:Accession: T43682  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1628 <WMA>  
 A:Cross-references: EMBL:AF05035; NID:95478223; PIDN:AAD43830.1; PID:G5478224  
 A:Experimental source: strain 972  
 C:GeneticS:  
 A:Gene: nup184  
 A:Introns: 6/2

Query Match	2.6%; Score 234; DB 2; Length 1628;
Best Local Similarity	19.2%; Pred. No. 3, 1e-07;
Matches 345;	Conservative 265; Mismatches 713; Indels 478; Gaps 76;
QY	34 SQIEAEINKWRRLLEG-----LSYKPP-----SPSSAEKRYANKDVAPLKEIGRISKFL 86
DB	43 SPVYGTILNSN-KTTEBEAKLSYDSHSHSKQSLSVK-----KISDIT 86
QY	87 GLDEQSVOLL---QCTLOEDYKTRDSVKTYLQDEQSQALIKADYYEERTC--IL 141
DB	87 GYNAQVAVYLVHVOYELNTQYPSQLDN-DSVLAQEFOR-----YYAEIISCKVKL 137
QY	142 RCVLHLTFYODERHPRVEYADCVDTLKEELVSKYQ-----179
DB	138 AFLQACTDADSKKH-----KMATRLIYSITQTAORSEMAQSTPISFCVRII 186
QY	180 -QFEELYKTAPEWETHGNMTERQVSRWFVQCIREQSMLEIIFLYAVAFEMAPSDLV 238
DB	187 DYLSKMTSQAPASPLFTENG---EBAISQWYFFHFNLLQQLLRVIFL-----STYSLVV 236
QY	239 LTRKFKQSGSGKQTNHLYDE-----TMDPVDRIGYFSAIIVEGMDIESLHKALDD 293
DB	237 CNSEMAISWFCMKKTYVLDHDEFMALDIDTSGMKETINVAIISINFISLEKQVLSF 296
QY	294 RRELHPADGLICQWDICMLTFPGDIPHPAPVLLAWALLRHTL--NPE-----ETSSVV 346
DB	297 KDNPSFPLSGNTIISLDMITQLSNSIGAASVLTGIALHLHLSNPDIPLIQNSVV 356
QY	347 -----RKIGSTAIQLNVFOYLRILQSLASGNDCTTSTACMCVYGLLSFVLT 394
DB	357 SSKILQNPQNSFQALILAAIKYDFPFLIHRILISL-----EDDVIQYGSKIMATLF 408
QY	395 SLEHLTGNQODIIDTCEVLADSLPELFWGERTSGGLIIDSVCGMPHILSPILQL 454
DB	409 SSAVSIVKESDSTMLCATTLFKTPQVQLF---ENSDSVRLNLPAPARPFESQVYL 465
QY	455 LRALVSGKSTAKKYVSPFLDKMSFYNELYK--HKPHDVISHEGDT---LWRQTPKLLYPL 509
DB	466 LIPFACLTQKQLVSSLSLHMTTFQSLPSGFAXEYELIPBNVYGNMLIELQESLHDSY 525
QY	510 G---GQTNLRIPQGTQGVN-LDDRAVLVMEVSYSSWTFLTCIEMLHNVSTADYIQH 565
DB	526 GFEPNNAERSLPKQTRIRIVSVTPYVVMWDIYSLMEAVGISLN---YIVRGLNSH 582
QY	566 QQRKPIIIDLVHKVISTLSIADCLPITSR-----LYMLQRLTVYVSPVDVIAS 617
DB	583 KSFVLTVLSSVPLFQTDVSGACELVHLASGDELDFINVIDLDLYFLSLVIEDAD 642
QY	618 ---CVNCLTVL-----AARNPAKWTDLRHTGFLPVAHPVSSLQWISAEQNAAGYG 668
DB	643 YQICVSSRLRLREFTRFAVD---VW-----AVYTRSLVCVSGSEKGIS---LE 684
QY	669 NLNMSRQPGEGVTTAFLRLT-----TLVKGQIGSISQSIGVPCVMVLKEML 719
DB	685 DVLPDYESINGVVDFTLAFDLYEILLDNCISVSVDPSIRLKTDFVYKAMFLCEVF 744
QY	720 PSYHKRKYNSHGVREOIGLIELIHALMCHETDHSHT-----761
DB	745 ANYLDMKTAIIQQYQIGHRFASILITKLAVNTGIEIFNPKTYNKKTLPRELSHYIQ 804
QY	762 -----PSILOFLICSLAYTEAQTYININGIVDTIDVMAAOPRSDGAEQ 808
DB	805 RFLVQDSNRNYLHPLLSVMDLINLVYDIPSTI-----SSPRAKAK-- 846
QY	809 GQGQLIKITYKLAFSVNNVI-----RLRPPSNVISPLQALSQGHAGNNLIAYLAKY 862
DB	847 ---MMILIS---SECAKTLICLRGFLNLRP-----SELRELFPSRSPDFN----- 886
QY	863 IYKHDPALPRL-----ATQLKRLATVAPM-----SYAACGNDAAAIRDAFLTRLOS 911
DB	887 -----CPRLLCCIAPIQLLSAL-TLAPWSEFETSLAAYMNSIDIYGRVCIOQLTN 938
QY	912 KIEDMRIKWILEFLTVAVE-TQPLGLEFLNLEVKDGSQSGSKFSLG-MWSCILHAV--- 966

Db 939 PIGSTNIEGSAVKFSLSSIMKGOQGLAVLFS-----GKPEPLDRMKSLHNVDVQ 989  
 Qy 967 -----LELISQOQDRWCPCPLHRAAIAFLHAIAMODRDSAMVLTKPFEMNLTS-- 1019  
 Db 990 LTKSKLSLAEKRLDSFINDILSQVPFEFISRNFWTSLGNLOEANFMNRIDAI 1049  
 Qy 1020 --PLFGTLSPSESEBSILETICALIMKIICLEY--VVKSLQSLKDT 1066  
 Db 1050 KLPLTWKLDGLSSVAQADLYILAAHATRTIAIQLHMKSLKNSKSIIDPLDSMKDL 1109  
 Qy 1067 LK-KFSIEKFPAYMSGYKSLA-----VHVAE----- 1092  
 Db 1110 VQNAFTIT---AYDSNHNALTRAFKHNGDLHISDLNLTGFLPLRGDNFYNIKLAKN 1166  
 Qy 1093 ----TEGSSC-----TSLLEYQ-MLVSAARMILLI---IATTHADIMLTSV 1132  
 Db 1167 MLNATEDTSFKISMMMSANENIISLDQAALLRSMISIFICAFVEFVXDATLPTLELKIM 1226  
 Qy 1133 R---ROLFLVDLD-----GTRALLVPASVNCRL---GSMKCTLLILLAQWK--- 1175  
 Db 1227 KWLKSLAEEDTIDVNVQELSAERPAALVFRIISQOTLAPISNVEKHLQSTLLTWRAIT 1286  
 Qy 1176 -RELGSYDEILG-----PLTEILEGVLOADOQLM-EKTRAKYFSAFIT-VLOMKEMKV 1225  
 Db 1287 TTKTSIYEDSGEMAYRRPLHVLVNTL---NRLSEKENLISVGFVSGLDLCHRK 1343  
 Qy 1226 SDIPQYQVLNVGCTLOEEVIALFDQTRISLALGATEDKDSMETDDCSRRRDQDG 1285  
 Db 1344 SOL--FEKAVINPTIEVYGVIVL-----NSLHK----- 1370  
 Qy 1286 VCVLGLLAKEL--CEVDEGDSWLQYTRRLPIPLITLLEVLAMKONHPTBATLHL 1343  
 Db 1371 -CYNSHLIRLOGLYISYINDSF-SVNDCL-----RLFSSHSLVLDGQYFPAALISF 1423  
 Qy 1344 LTLTAQTQCATAVAGAGITQISCLPLISVYQLSTNGTAQTPSARSLDAPSPGYRL 1403  
 Db 1424 LLISSSPAGAEQIVMNGFFVSIWESPLST-ALSTGGIGLDGS---SIQKIN---IRG 1475  
 Qy 1404 SMSMEQLKTLRNPLPEALDFVGHQERTLOCLNAVTVQ-----SLACL 1450  
 Db 1476 ILPLFLFVFKLGRINNDMEFY-----LLAFPOIQVALLMWCOPSSISLASI 1525  
 Qy 1451 EEAHTVGTILQLSNPKMWHFHLPOLMRDIOVNLGVLCQCTSLHSRKKLQHYLKN 1510  
 Db 1526 DES---FMVLVLPDLQGFN---PALQETR-----LAELKIMLGIDYLISHPNFLS 1573  
 Qy 1511 GDGLPSAVACQVRPPSNAASAPSSSKQPADTEASEQALHTVQYGLIKLISKTLAAR 1570  
 Db 1574 SLAIPALYEGEVAIDVIGIKELSDQLTKETTSQPAKIRIQ-----LDLRLALLE 1627  
 Qy 1571 H 1571  
 Db 1628 H 1628  
 RESULT 2  
 T06075  
 hypothetical protein T9A14.40 - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 09-Jul-2004  
 R/Accession: T06075  
 R/Author: M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Newes, H.W.; Meyer, K.F.X  
 submitted to the Protein Sequence Database, March 1999  
 A/Accession: T06075  
 A/Reference number: 215184  
 A/Molecule type: DNA  
 A/Reidues: 1-561 <BEV>  
 A/Cross-references: UNIPROT:Q97014; EMBL:AL035656; GSFPB:GN00062; ATSP:T9A14.40  
 A/Experimental source: cultivar Columbia; BAC clone T9A14  
 A/Genetic: A:Gene: ATSP:T9A14.40  
 A/Map position: 4

A/Intons: 9/3; 45/3; 64/1; 104/2; 149/3; 202/2; 258/3; 291/3; 317/3; 337/3; 390/2; 457/-  
 C/Superfamily: Arabidopsis thaliana hypothetical protein T9A14.40  
 Query Match 2.44; Score 215.5; DB 2; Length 561;  
 Best Local Similarity 22.28; Pred. No. 1e-06;  
 Matches 132; Conservative 99; Mismatches 244; Indels 119; Gaps 24;  
 Qy 86 LGLDEBSVOLLOCYLOEDRGTRDSYKTVLODE-----ROSQALIL 127  
 Db 9 MNLDEISYILVERSMQOEY-GTTDSVAQELTQEFIMSGFLHINQCFVWVWESCEFL 67  
 Qy 128 KIADYYEERTCILRCVLLHLLTYFQDERHPRYEYADCVDLKEKELVSKYRQGFELKYT 187  
 Db 68 TLISLFS--NVILEDASVLTAPVPTKYKVAHSHTCI-----ICSVAPRESSIKE 119  
 Qy 188 EAPWETHGMLMERQVSRWFVQCLRQSMLEIFLYAYFEMABDVLVTMFEQ 247  
 Db 120 EAVVLISDG---LER-----ROSSVLEDLISCF-----PRMGILS----- 153  
 Qy 248 FGSRQTRHLVDETMDPFVDRIQVFSALILVEGWDISLHKCALDDBRELHQFADQGL-- 305  
 Db 154 -GSYNSKLAIVSVAQHSACRVOIQMLMILITIDMEN-----LLQVHVDGVP 201  
 Qy 306 ---IC-----QPMDCIM--LTFGDIPIHAPVLLAVAL---LHRTINPETSVAVKIG 350  
 Db 202 RSGTCVPSIYDVQEMDATISSLNTSEVNAAGPLVLAVALCLISLPGKEESPFLMDID 261  
 Qy 351 GTAIQLNVQ----YTRILQSLASGANDCTTSTACVYGLISFVLSLEHTL---G 402  
 Db 262 HVSYVHQAFEAASLSYFLEILQSLMLNDPDGPISGHRSVTRFISAFIASEYIMQLQEDG 321  
 Qy 403 NQODIITDAGEVL-ADSLPELFWGTE-----PTSGIILDSVCGMFLPLSLDOLR 456  
 Db 322 TLIELDILISKVGGESLCCQFMDKSFYDGPRLCLPDLFS---EPFSAFIFILS 378  
 Qy 457 ALVSGSTAKVYSFLDKMSFYNELYKHPDIVISHE--DGLWRRQTPKLYPLAGQTN 514  
 Db 379 SLSEGSMPAEVGVYFLDKSGVSTLF-----DILSDPADASQVLETSLRH-IPGLEG 432  
 Qy 515 LRIQGVQGM--LDRAVLVWRWYSYSSWTLFTGCEIMLHVAVSADVYQHQRKPI 572  
 Db 433 LVIPSNTRGRILRVISNTVLYRWVYSLSGIIVLIRLANKLYGNRRAP-----VT 485  
 Qy 573 IDLVHXYISDLSADCLPITSRIYMLQRLTVISPP---VDVIASCVNCLE 623  
 Db 486 DELRRVTRFKAVCSFLNLSHFYQESTVNGKMSDVRAVDIDICNSVRSLE 539  
 RESULT 3  
 S47446  
 nucleoporin Np188 - yeast (Saccharomyces cerevisiae)  
 N/Alternate names: protein YML103c  
 C/Species: Saccharomyces cerevisiae  
 C/Date: 02-Dec-1994 #sequence\_revision 02-Dec-1994 #text\_change 09-Jul-2004  
 R/Accession: S47446; S61617  
 R/Author: R.; Barrett, B.G.  
 submitted to the EMBL Data Library, August 1994  
 A/Accession number: S47445  
 A/Reference number: S47446  
 A/Molecule type: DNA  
 A/Reidues: 1-1655 <BAR>  
 A/Cross-references: UNIPROT:P52593; EMBL:X80835; NID:G530339; PID:G530341; MIPS:YML103C  
 R/Zabel, U.; Doye, V.; Tekotte, H.; Wepl, R.; Hurt, E.C.  
 submitted to the EMBL Data Library, August 1995  
 A/Accession number: S61617  
 A/Molecule type: DNA  
 A/Reidues: 1-1655 <ZAB>  
 A/Cross-references: EMBL:X90580; NID:G1150597; PID:G1150598  
 A/Genetic: A:Gene: SGD:NUP188  
 A/Cross-references: SGD:S0004571; MIPS:YML103C  
 A/Map position: 13L



## RESULT 4

T41267

hypotheetical protein SPCC290.03c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: T41267

R:Lyne, M.; Rajandrem, M.A.; Barrell, B.G.; Gilbert, H.; Lauber, J.; Duesterhoeft, A.

submitted to the EMBL Data Library, January 1999

A:Reference number: Z21982

A:Accession: T41267

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1647 (LYN&gt;

A:Cross-references: UNIPROT:P78847; EMBL:AL035260; PDB:CAA22873.1; GSPDB:GN00068; SPDB:

A:Experimental source: strain 972h; cosmid C290

C:Genetic:

A:Gene: SPDB:SPCC290.03c

A:Map position: 3

A:Introns: 1562/1

Query Match 2.1k; Score 192.5; DB 2; Length 1647;

Best Local Similarity 17.2k; Pred. No. 0.00024;

Matches 343; Conservative 305; Mismatches 673; Indels 677; Gaps 84;

45 RLLEGLSYKK-----PPSSSAEKYKANK-----DVASPLKEGLRISKFL 86  
31 RLILICLESYKDFLKLADPANANSRKLEGEVELGVYINVEQFQLSLTSTQL 90  
87 GLDEEGVOLLQCYLQEDYGRTRDSVTVLQDEROSQALIKIADYYEERTCILRCV-- 144  
91 NLDEICQASLLQRIEASQNLDRTPVQALY-----PFLAREQLLECLES 136  
145 -----LHLTYFQDERHPYREYVADCVKLEKEVSKTRQGFELYTEA 189  
137 LTRVWGLKDESDISTALSKYLSLQICENNNLVCTCID-----TPIIDNVSEILKSEA 191  
190 PTWETHGNLWTEROVSN--FVOCLEBOSMLEIIFLYAVFEMAPSDLL-----VLTK 241  
192 -----GGQILGVSEVDPQEFIR--LSHBAVALETTISVILYQLAKVDLFONSHFESLLV 245  
242 MFKEQSGSRQTNHLYDETMDPFVDRIGYFSALILVEGMD-----IESLHK 288  
246 MLRK--YDSPKNAVLPLTYAFIDKY-----LEVBYLPDQKVQRLSNSVELLQKHQ 297  
289 CALDD-----RRELHQPADGLI 306  
298 AIIQSPQDWSSQFKNIIIGIMVTRLNATKQIEKVPFIDYETTIKNANEIIQNGVF 357  
307 CQMDICMLTFQD-----IPHHAPVLLMALLRHTLN-----PEETSSV 345  
358 SDMTLLVYPRQSETEGEMWAFKRSRITVNMSLIRPTIASITFSELSRFAQAVSY 417  
346 VRKIGTAIQLVNQYTLRLQSLASGNDCTTSTACVCYGLSFVLTSLHETLGNQ 405  
418 MPDILKTLRLLEBRVLTNTTPTSPISPEQIEEFPPPEEYVLLSSYTY-----NVS 470  
406 DIIDTACEVADPBLPELFMGTEPTSGIGITLDSVCGMFPHLSPLQLPALVSGSTA 465  
471 WISD-----FMDIESDMVGFLTWSMGSOIPGIIATFILLASLAKNTTSA 516  
466 KKVVSFLDKMSFVNELYKHKPHVDIASHEDGLMRQRPKLILYRGGQNLRIPO--GTVGQ 524  
517 SKIT-----ELFSEP-----IPEVGHDES 535  
525 VMLDDRVLVWEXSWTLFTCEIEMLLHVSTADVIQ--HCOR-----VPRPI 573  
536 LMTISPS-----WSYIFNVFRYIISHLKPVQVTVSSGLARVHTDPSLIDSDALILQAYI 591  
574 DLYHKVISTDLISND--C-----LIPITSRIYMLIQLRLTVTISPVDVYASCVNCLTVLAA 627  
592 LIFSSVVRQDAQIATSTECENDLNPIATLFELECRLL-----PDSYRICIVRALESIAH 645  
628 RNPAK-----VWTDLRH-----TGFLRFAVHPVS--SLSQMTISABGMNNGVGNL 670

646 LSTGFSNNALMTALDMNFVSVLFDVQGLAPSPISAKSLTKPVTSCGPLNNIRRL 705  
671 LANSQPOGEGYVTTIAFLILTLVKKQ-----LG--STQSGVLPCMFPLKE 717  
706 TVNLE-----KMSIFVNLTLTRNKSLELVNLPENMIGASYRPPQPYDVYVET 758  
718 MDSYHKWY--NSHGREQIGCLILELHAIIIN-----LCHETDHSHSTPSIQ----- 765  
759 FVASSTWRMLRMDGIRLOYAC--LQYMLAVDGLNIDILYSRISSKYRDLQNNNL 816  
766 -----FLICSLATTEAGOT--VINIMGVDTIDMVAAPRSDGAGQGGQGLILIK 816  
817 HVTLTRHPALSLLEALVTESVYGLPDLVYVGFQLEDD--SVPKT-----IVI 863  
817 TVKLAFSVTNVRLKRP--SNVY-----SPELQ 843  
864 TVSASLCLRNNAVLSQVLFKNVVPYIAELGISYILDTISRRAKVEVMTIRISIVHL 923  
844 AL---SQGHAGNNLAVL-----AKYIYKHDPALPRLAIOLKRLATVAPMSVYACLG 895  
924 ALVGSRHKCFPLSALISLYLDABGFNNKRPDENKLCSTII--RTANSKRILIFGI 981  
896 NDAATIRDAFLTRLOSKEIDMRKIMLEFLTYAVETOPGLIEFL-----N 942  
982 ---RTFESQFLTLSTNDESS---LILKILLNNLKSGGVYSLALLILGPDISTNVIT 1034  
943 LEVKGSDGSKFSLGMSCLHAVALLEIDSQ--ODRYWCPLLRHAA---IAPHLAMOD 998  
1035 LRDPQGVYGSRV-----SLMNSLIDPIEGRTIYNGIEMTPVIMVQALEIYAFICSC--PL 1087  
999 RDSAMLVATKP-----KFWEN---LTSPLFGTSPSESEBSILETC---ALINKII 1047  
1088 TSEVTLSTVIRARPELVKMWQGEPILOQVYQNGGFS--SEBSWASRCIRSRTOIANNL 1146  
1048 CLEIYVYVKSLSPOSILKDTLKKEFSIEKRFAPWSGVYSLAVHAETECSSTLLEYOML 1107  
1147 ATEIHVA--ASVQGN-----KLYNEVAST---IRTNKHTSTELSGQE-- 1185  
1108 VSAMRLIITATTHADIMHLDTSVRROLPLVDLGTKALLLVPAVNCRLSGSKCTLL 1167  
1186 -SGFKL-----EFMDI-----LRIDPOSIT--FELPMIPGFNL 1216  
1168 LILRLQKRELGSVDLELGPTEITLEGVLOADQOMKTKAKYPSAITYLQMKEMKVS 1227  
1217 NMFTRFDR--GNSD-----FOPTERVILKIRLFEAEAT 1250  
1228 IPOYSQVLNVWCETLQAEVIALFPQTRHSIALGATKDSMETDQCSRRHRDQRDGVC 1287  
1251 -----SLGSAEKSYSWDEQNSKJELAQR----- 1276  
1288 VLGHLAKELCEVEDG--DSWLQVTRRLPILPTLLTEVLSLMKONLHPTTEATHLL 1345  
1277 -LNTFNHVVLLQDIGHCLTAMARLTGIL-----VDCNEVISDVHFDPIWGLV 1325  
1346 TLARTQCATAVAGAGTGGICLPLBSVYQSTNGTQOTSASAKSLDAISWPEVYLSM 1405  
1326 -----RLVLPQVTVNHLGQGT-----V 1343  
1406 SLMEQLKTLRYNPLPALDFVGVHQBRTLOCLNAVTVOSLACLEADHTVGFILQSLN 1465  
1344 SVTSSVLETT---LPHALKRIGALKPEBLG-----KSTY 1374  
1466 FMKEWHFLDLQMRDIGNLGYLCOACT-----SLHSRRMLQHYLQNNKGDGLPS 1516  
1375 FMEGIDHIVIGLAKGIC-----CQGSDESIRENLYGSLIILVFPQKHSANGDKYDP 1428  
1517 AVAKRVPRPSASAASSSKQPAADYBAEQALHTVQYGLKILSKITLALAHFTPRDV 1576  
1429 VFQDAFOK-----LITPNLLTSQFFDVLTKD--ALYT--NCSWELSVIILNPLHNVSDI 1480  
1577 COLLDQSLDAEYNFLFALSTPTPDSVAPSGTLLATVVVALMMLGELDKKKEPLT 1636

Db 1481 STHL-----YKYLRRNFVSSFLDA-FSRAFSLLSSNKDVLVLSGLEAGQ----- 1526  
 Qy 1637 QAVGLSTQAGSTRTLLSLMTMENCFYLLISQ-----AMRYLRDPVHPDCKRMK--- 1688  
 Db 1527 --CLITFAQKMLVSHVLTIDYIKLAWQMLQCKGGIYLYLPV-----ORLMIQL 1578  
 Qy 1689 -----OELSELSTLLSLSRFRGAPSSPATGVLPSPQCKSTLSKASPE 1735  
 Db 1579 LQPIVLVMTVLTKEISNKDILLQSIPLSRK-----LQDSVQTGADQ 1622  
 Qy 1736 SQEPLIQVCAFRHMQR 1753  
 Db 1623 ALSEPNTVKKYVETISR 1640  
  
 RESULT 5  
 S56811  
 probable membrane protein YJL039c - yeast (Saccharomyces cerevisiae)  
 N/Alternate names: hypothetical protein J1216  
 C/Species: Saccharomyces cerevisiae  
 C/Date: 08-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 09-Jul-2004  
 C/Accession: S56811  
 R/Poln, T.M.; Aljinovic, G.  
 submitted to the Protein Sequence Database, September 1995  
 A/Reference number: S56793  
 A/Accession: S56811  
 A/Molecule type: DNA  
 A/Residues: 1-1683 <NOV>  
 A/Cross-references: UNIPROT:P47054; EMBL:Z49314; NID:g1008164; PID:g1008165; GSPDB:GN000  
 A/Genetic: A:Gene: SGD:NUP192; MIPS:YJL039c  
 A/Cross-references: SGD:S0003576  
 A/Map position: 10L  
 C/Superfamily: Saccharomyces cerevisiae probable membrane protein YJL039c  
 C/Keywords: transmembrane protein  
  
 Query Match 2.1%; Score 192.5; DB 2; Length 1683;  
 Best Local Similarity 18.2%; Pred. No. 0.00025;  
 Matches 307; Conservative 239; Mismatches 553; Indels 589; Gaps 81;  
  
 Qy 2 IRKSKITSVLSF---C-RSSRELTILLGSALELSQIEALELAKMKRRLLEGSLTYKPP 57  
 Db 108 LRRQYIIQIVSFIVNCHDELTLYOELIKGAL--VSNI----- 144  
 Qy 58 SPSSAEKVKANKDVASPLKELGLRISKPLGDE-----EGSVLLQCLQEDYGRTRDSV 112  
 Db 145 -----LSAFKFIHTQLSEIKQOINKAQILBNYNAIPQONIKRRPFLAREY---DIL 193  
 Qy 113 KTVLQDEROSQALLIKIADYYEERTCILRCVHLITVYFODERHPYEVAYADCVDKLEKE 172  
 Db 194 SQLLYG-LVDRKGAIMKKNDF-----ILSLHH-----VSELSN 226  
 Qy 173 --LVSKTRQRFEEELKTEAPFWETHGNIMTERQVSRNFVQCLRQSMI-----LEIFLY 225  
 Db 227 DFFLIYTPPAFFHLFASLRVLPDADVKLHLSQ---FMKDKKDSITYRKVKYALFIF 281  
 Qy 226 YAYEMAPSDLLVLTAKFKEQFGSGRQTRHL-----VDETMDFVDRIGVFSALL 277  
 Db 282 FAYF-----IGCKEDPKRADMTDKTVDDEMTSAVE-LGALIQOLI 324  
 Qy 278 VEGMDIESLHKALDRRELHQFADGLICQMDCLMTFGDI---PHHAPVLLAWALL 333  
 Db 325 -----PAADTSIV-EQDKSMELFYDISLIERHPRILPIKQL 361  
 Qy 334 -----RHTLPERTSSVVRKIG-----GTA-----IQANVFQYLTR 364  
 Db 362 DDEKIFSGQTNNSTNPAATDNMGRGLMNPSPGMMSTGTALNSMPNNVVEYSTTI 421  
 Qy 365 LL-----QSLASGNDCTSTPACMCVYGLASFVLTSL---ELHITLGNQODIITPACSVLA 416  
 Db 422 VLSQGTQEFELSSPDDVQITITDC---AFLITKIKDAEEDSLSGEDITLDDISLKA 476  
 Qy 417 D-----PSLPE--LFWGTEPTSGIGII-LDSVCGMFPHLSPILLQLLRALVS 460

Db 477 DLRRFLSYFYASRPREYSCFTFMSDKNASNAFIEWCSRCN--DNLMRSCFYLMVSSLS 534  
 Qy 461 -GKSTAKVYSFL---DKMSFYN-----ELYKPKPHVISHDEGTLMRQTPKLLYPLGG 511  
 Db 535 FGPEMALNIVHYHGENSSISMKNIACLSDYTK---ISNFSSLSHKRQ----- 580  
 Qy 512 QTNLRIPQGTVGQVMDDRAYLVLRWEYSYSWTLFTCEIMLLHVSTDAVIGHQVRKP 571  
 Db 581 -----QPSSESTHN-----DIDSTAAVALEGLNEBA 605  
 Qy 572 II--DIVHKVISTDLSIADCLLPITSRIYM-LLQRLTVIISPVDVIASCVNCLTVLAA 627  
 Db 606 VIFLSLLTLVGSVTVQVDEVDKSSLSKVPSDVLFEFTKINTPLVGAAFKVYSNLVPEKLE 665  
 Qy 628 RNPAAKWTDLRHGFLPFPVHAPVSSLSQMSAAGMAAGGONLLMNSQOQEGVGTIAF 687  
 Db 666 SSRTEKWS-----FDLSLIFKDSISLN--YSSBSYR-NAFTNVL-----KYSVGLGF 709  
 Qy 688 LRLLITLVKQQLSGTOSQGLVPCVMPFLKEMLPSTYHK--RRY----- 727  
 Db 710 LQLFPHLISHSHENNSYVWPGKLAEPTRLGGYRKVGIMPFYDFINDILAHVDQIVD 769  
 Qy 728 --NSHGVRQOI-----GCLILELHAIALNL-----CHETDLSSHSTPSIQFLC 768  
 Db 770 IRMKRAVQPLIKITITGLCSFDYSYLNSIPAAANLDALVDCENFPNVYQECRAP--- 826  
 Qy 769 ICSLIATYE-AGQVIVINIMGVDTIDVMAAGRSDBAGQGGQGLIKTKVLAFASTNN 827  
 Db 827 IFNYIFEFKIKYKSIFFNVADVGVQDLSIELEG-----GKQAEELL---QLAVKLINK 874  
 Qy 828 VIRLKPPSNVVSPEQALSQHG-----AHGNM-----LIYVLAITYH 865  
 Db 875 VLVQYQ--ETVVEELFPYKKGKTDYFELPKNSYLHGARSYDAIFNIPVLAHGLYV-G 931  
 Qy 866 KNDPALPRALIQILKRIA---TVAPMS-----VYACLGNDAAIRDAFLRLQSKI 913  
 Db 932 VDDQIATNSLRILAKLSERSNGSVASLSKRNKLITFDSV-DESAKIKDAFIQLESSI 950  
 Qy 914 ED--WKIKMILEPLTVANETOP---GLIELFLNLVKKQSGSKSEFSG----- 958  
 Db 991 TDAGVALKIKELIDFLTSNLSNYSRTWTISHLLGQV-----SNVISGPNLAFTISS 1044  
 Qy 959 -----MMSCLAVVLELDSQODRYWCPLLRPAIAIAFLHALWODRRDSAMLV-LRTK 1010  
 Db 1045 GTSLDLSLISVLASINSITKNID--YAPMRLATATALEIILKCNPNLSGLYSYLIK 1102  
 Qy 1011 PKFWENLTSPLFGTSLSPSETSEPSLLETICALIMKIIIEIYVYVKGSL-IDOSIKDTLKK 1069  
 Db 1103 ENFERIRM-----ILDPOVTR-----FTWNGSPEDNSTERCKN 1137  
 Qy 1070 FSTIEKRP-----YMSGVYKSLAVHVAETGSSGTSI----- 1101  
 Db 1138 F-IESBSVGAFSLAYRNWTQYL--GLFTHKISFGSTSEVLTVYNYLISNTMYSVRLF 1195  
 Qy 1102 -----LEYQMLVSAEMMLIATTHADIMMLT-----D 1129  
 Db 1196 SFPLDPLNTGNICPKKELISFTNVPLNLEQVTLINKCSGNIYDFHKMENMLRIKRYBAE 1255  
 Qy 1130 SVVRQQLFDVLDGTALLLVPASVNCRLRGSKCTLL-----LILLQWKELGS 1180  
 Db 1256 SLHSNFSFLTV--SKQFQKADAVECIRAKSHFTNIIISHKALEINLSYLAHVMVQ---- 1308  
 Qy 1181 VDDELIGPLTEILGCVLAQADQQLMEKTKKAK-VFAFAITVLAQMKMKVSDIPQYSQVILNV 1239  
 Db 1309 -----LVQIITVQKLEPSTRSNFLEVFGLIP---KISD---YIEFNFTFS 1350  
 Qy 1240 ETLQOEVIALLD--QTRHSIATGASATD-----KDSMETDDCSRSR 1278  
 Db 1351 EBLVSLAVLFLFDLYNDRKILTDKGVLDGRLYQLFKTCIGINSPLSSVALRSDFYLLAN 1410  
 Qy 1279 H-----RDQRDG-----VCVLGLHIAKELCEVEDDG 1304



Db 1411 HYLRSVLSDQVSEKVLQDLRLSGSKVLVEIIMNDVYGEGRVYTGILLDLSLIQLANRS 1470  
Qy 1305 -----DSMQLVTERLPILPTLLTLEVSLSMKONLHFE-----ATLHILLTLAR 1349  
Db 1471 KENFIDLSMKTTLTLIRLKNLTDALNSTTHINIDLLYELTAKFATVFLINVAE 1530  
Qy 1350 TOOGATAV 1357  
Db 1531 TRGASAL 1538

RESULT 6  
T37919  
GCN1 homolog - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T37919  
R:Conor, R.; Churcher, C.M.; Bartell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: 221754  
A:Accession: T37919  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2670 <CON>  
A:Cross-references: UNIPROT:Q10105; EMBL:Z68198; PIDN:CAA92385.1; GSPDB:GN00066; SPDB:SH  
A:Experimental source: strain 972h-; cosmid c1866  
C:Genetic8  
A:Gene: SPDB:SPAC1866.05c  
A:Map position: 1  
A:Introns: 50/1

Query Match 2.04; Score 184.5; DB 2; Length 2670;  
Best Local Similarity 18.64; Pred. No. 0.0019; Matches 697; Indels 699; Gaps 89;  
Matches 380; Conservative 269; Mismatches 697; Indels 699; Gaps 89;

Qy 20 LMTLLGRSALRELSQIEAEI-NQHWRLLEGLSYKPPSPSAEK--VKANKDVASPL 75  
Db 797 LMDVLEKSSKLDKMTKDYETKRWAEVAKQSAKKPALKSDQALVDAQDAEAKI 856  
Qy 76 K-----ELGIRSKFLGLDEBQSVQLQCYLED-----YKGTDSYKTV 115  
Db 857 RSRVNLALSLERGLGIIRSLG---BAVQLAPALWEDADIVLLFHNVLKYSESPFKNL 912  
Qy 116 LQDRQSOALILKADVYEEER-----TCILRCYLHLTYFQDERHPYREYVADCYDKL 169  
Db 913 AYD-----TFLLTLKASGFERLGDRTSYSSLSLIL-----AHTFSVNSSENTEL 958  
Qy 170 EKEIVSKYRQFEELYTEAPTEWTHGNLMTERQVSRWFVQCLREQSMLEIIF--LYYA 227  
Db 959 TKSILYLRFAIEQNY-----FEPQMFACIFPLLYDL 990  
Qy 228 YFEMAPD-----LVLTYMFKEGFGSGRQNR-----HLVDETMDPVDRIQ 270  
Db 991 TFNITNSDEDEDEALQLVTEILEFQALYSASLRMSKLIKSLHLEIAPTOQENKN 1050  
Qy 271 YFSALLIVEGMDISLHKCALDDRRELLHQFADGLIQDMDCMLTREDIPIHNAFVLLAW 330  
Db 1051 -SLSLICEG-----LHSTYDEELNL----- 1070

Qy 331 ALHTHTNPEET--SSVRIKGTALQIANVQYLRLLQSL-----ASGNDCTSTA 381  
Db 1071 -LISNLHPBSISAVLQAL--QAFDLSREPIKEIETLEYXDNETNAGIAHQISTQN- 1126  
Qy 382 CMCVGLISFVLTSLHTLGNQ-----QDIITDACEVLAD----- 417  
Db 1127 -----GLDATERSEFELQIFPTQSDVYQIIGKSLDLDLDEFELQGFIRKELMRYRE 1181  
Qy 418 ---PSLEEL-FMGTEPISGL-----IILDSVCGMFPH-----LISPLQLL----- 455  
Db 1182 NALPSAPYDEYGIKKEITGRDLGRJARBSVAVSFFHISKYSLSNLLPLLELLTASE 1241  
Qy 456 -RALVSGKSTAKKYSPF--DKMSFYN-----ELYKHPHVIDISHEDSTLWRQ 501

Db 1242 AEAQIPVTDASQKVSSEKLEAGKLAIFQSGAHQVEALMELEOKLNVDSLPTDANLELR 1301  
Qy 502 TPKLLEYLGGQTNLRIPQGTGVGM--LDDRVLVREYSSWTLPTCEIEMLHVST 559  
Db 1302 ATVVLV-----GTVAQHLPSNDPLAV-----MDSLSVLSLT 1334  
Qy 560 ADVLQHCQVKEPIIDLVHKVISTDLSIADCLPITSRIYMLQRLTTVISPPVDVIASCV 619  
Db 1335 PSE-----SVGLAVAVCLPPLVKK----- 1353  
Qy 620 NCLTVLAARNPAKXWTDLRHTGFLPFVAHPVSSLSQMSISAEGMNAGYGLNLSN-----E 675  
Db 1354 -----SLGSKSEYVEL-----LSNKNLSTSLAD 1377  
Qy 676 QPQGEVGTIAFLRLITTLTKGQSGTQSGVPCWFVLEKMLPSYHKRKYNSHGREQ 735  
Db 1378 QKGAAY-----LAGLVKG-----YGIKAFQDFNLIDLSLISNRQNA--THRQ 1420  
Qy 736 IGCILIELHAIILMCHETDLHSHHTPSLOFLCICSLAYTEAQTVINMGIGVDIYDMV 795  
Db 1421 VALFAVAFSRILGITYE-----PYLPDL-----LPILLTSFGDNANEVREATMDAVKQI 1470  
Qy 796 MAA-----QPSDGAEGQ-----QGOLLIKTV--KLAFS 823  
Db 1471 MSQLSAFGVKLLPTLLDGLNEYVWRSKASVEILGMSYMAKQSLVFLPTIIPKLSEV 1530  
Qy 824 VTNNVITLKEPNS-----VPSLEQALQHGAGNNLAVLAKIYI-H 865  
Db 1531 LTDHSGQVNTANKSLRFGDVISNPEIQTLVPLTLALSDCTRYTDALALKTSPVH 1590  
Qy 866 KHPDLPRLAIQLKRLATVAPMSVYACLGNDAAIIRDAFLTRLOSRIEDMRIKVMILEF 925  
Db 1591 YLDP--PSLL-----VITPIKY-----GLRENAQTKQSAKIRPGMASLTPEN 1634  
Qy 926 LTYAVET-QPGLIELFLNLSVKD-----GS-----DSKSESL--GMSCLHAYLE 968  
Db 1635 LAVVLESIMPLRLREVLD--FVPDTRATAKALGSLIEKKEKFPPLIPELFLVLRSECS 1693  
Qy 969 LIDSQ-----QODRYWCPELHRAIALFLHLMQDRDSAMVLATKP 1011  
Db 1694 EVDRQGAAGUSETIAGLGLARLED--VLPEILKNTSSPVPHI---RSSFISLLYLP 1746  
Qy 1012 KFMENITSLPFLGTSLPP-----SRTSEPSILETCALIM-----KIICLEIYVYVKS 1058  
Db 1747 ATGSSRFPQPIAARIPILISGLADDSLVQTAIRAKAMVNNYATGSVLLLEBELKGL 1806  
Qy 1059 LDQSLQDTLKKFSIEKRFAYMSGVYKSLAVHVAETEGSCSTSLLEYOMLVASMRMLIIA 1118  
Db 1807 FDNAMRIRLSVQVGDVLEKLAGINKKALOEDDEEGTHSD-----VSRKALDIIIG 1859  
Qy 1119 TTHADIMHLDSVVRQLFDVLDGTALLVPAVNVCLALGSMKCTLLILLRQWREL 1178  
Db 1860 QERHDIRLSTLYYRQDI-----AAVVRTPA-----IQIWAIV 1893  
Qy 1179 -----GSYDEITLGPTEILEGLQADQ-----QIMEKTKAYSAFIVLQ--M 1220  
Db 1894 VMTPTVREILPITLITISVNSLSSNDKRTMCYKSGIDLLKKAQFVYLQQLPVLQGL 1953  
Qy 1221 KEMKVSQIPQYSQVLTAVNVCETLQEVIALFPDQTHSLALGSATEDKSMETDSCSRSHR 1280  
Db 1954 ESANGD-----RIGVCIALEB-----LINSATPEQLEIYSDDPVAVVR 1993  
Qy 1281 DQDRGVVGLHAKELCEVDEGDSVLQYTRRLPIPLITLLTLEVLBRKQNLHFEAT 1340  
Db 1994 ALMDG-----DLEVARETAEAFFDSLOSITGDAVVDVLPOLLKLESNQSQALSALREI 2049  
Qy 1341 LH-----LALTL-----ARTQCATVAGAGITQSICTPLSVYQSLSTGTA 1382  
Db 2050 ISRSSTIFPVLYITLLIKKYVSAPNABALSSLAQVAVTINKLPSTILNLMBSLASTG 2109  
Qy 1383 QTPSASRKSILDAPSPGVYRLSNLSME-----QLLKTLYNPLPALDPVGVHQRITQL 1438  
Db 2110 DDIVALNGAID-----KVNLSVKKQEGIQIIMAHFYSF--SESEDF-----RKR----- 2151



```

QY 1439 NAVRTVQSLACLEADHTVGF-----LQLSNFMKEMHFLPQIMDIQVNLGYLQACTS 1494
Db 2152 -----LFAAEHMLVFQONCKLDYRYRGDWMHFTTLTEDKSQDVVAVAQAQN 2200
QY 1495 LIHS--RK-----MLQHYLONKNGDG--LPS-AVAQRY-----QR 1524
Db 2201 TLVSALRKQQLDLSVSIAYHSLRDVSGQVNLPAPEVAQGVNSLIPFLYGLMGTMDQR 2260
QY 1525 PPSAASAPSSSKOPADTEASEQALMT-----VQYGLKLIS 1563
Db 2261 EOSALGIDIVLK-----TEPSKLPFYQTGTPLIRIGERFPVEVKCALTYLNTIIS 2315
QY 1564 KTLAALRHFTPDVQCILLDQSLDLAEVNFALPFLSTPTPD---SEVAPSGTLATVNV 1620
Db 2316 KISTLRFPFLQOLR-----TFKCLGDPSSSEVISRAATLGTITIT----- 2357
QY 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRTLKSLMTWENCFFYLLISQAMRYLRDPAVH 1680
Db 2358 -----LQRLAPL-----ITELVSGARTPDAGVRKAMNALPAVVSQGNNEASAE 2405
QY 1681 PRDKORMQELSELSTLSSLSRYFRGAPSSPATVGLPSPQKSTSLSKA--SPESQEP 1739
Db 2406 A--IEQLDEISAESSEHMTCAKY--GALFSH---LPDAQAKQLLESKVLSLETQSE 2457
QY 1740 LIQTV 1744
Db 2458 FSVLI 2462

RESULT 7
T06074
hypochemical protein T9A14.30 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
C/Accession: T06074
R/Beyan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
Submitted to the Protein Sequence Database, March 1999
A/Reference number: 215184
A/Accession: T06074
A/Molecule type: DNA
A/Residues: 1-1073 <BEV>
A/Cross-references: UNIPROT:Q9T013; EMBL:AL035656; GSPDB:GN00062; ATSP:T9A14.30
A/Experimental source: cultivar Columbia; BAC clone T9A14
C/Genetic8:
A/Gene: ATSP:T9A14.30
A/Map position: 4
A/Intons: 45/3; 93/3; 129/3; 184/2; 343/1; 451/3; 977/3; 1047/1
C/Superfamily: Arabidopsis thaliana hypochemical protein T9A14.30

Query Match 1.9%; Score 175.5; DB 2; Length 1073;
Best Local Similarity 17.6%; Pred. No. 0.0018;
Matches 213; Conservative 189; Mismatches 416; Indels 393; Gaps 51;

QY 694 LVKQGLSTQGLVPCVMFLKEMLSYHKMRNNGVREQLCLLELHAIIN----- 749
Db 3 LVBSGL---ENDVVFALVPSLOYITLASHYWKYKNGHGMKVKTLKVIETLMKTCLEFSKF 59
QY 750 -----LCHETDIHS-----SHTPSLOFLICIS-----LAYTEAGQT-VIN 783
Db 60 STKLRLVLDLILINDASVHSLFRIICTTQNLLETIOVCAQVLSKLFALAESQLYIIS 119
QY 784 IMGIGVDT-----IDVMAAOPRSDGAEQGGQGLIKITVCLAPSVTNNVIRLKPSSNV 838
Db 120 NAFEGLDNKKQPALVLAIFDEDESDSN-----VQSKRKDASSIPDMACKSRLHTTIL 172
QY 839 SPLEOALSQGAHGNMNLIAVLAHYKHDPALPRLAIOQLKRLATVAPMSVYACLGND 896
Db 173 QYERATDPFDRHTDILGLL-----DFLKTLMQEA--GYA----- 207
QY 899 AAIRDAFLTRLQSKIEDMRI-KVMILEFLTVAVEQGLIELFLNLEVKD---GSDGSK 954
Db 208 -----NMLEFFKASKKULMQEFSDIISQAS-----KINDSTVGSIGKEE 245

```

```

QY 955 PS--LGWMSCLHAVLEILDSQQODRYWCPPLHRAAIAFLHALMDDRDSAMLVLRTPK 1012
Db 246 ISKLIVYQQAASVLEIM-----ACNNFLYKGL 274
QY 1013 FWNELTSPPLGT-----LSPSET-----SILETCALINKITCL-- 1049
Db 275 FASLKKPCVETKKTANGVSPKLTWTADSDPKDIFSKWCDISVD--GIQVSSLDG 332
QY 1050 --EIVYVKG-----SLDQSLDKTKFSIEKRFAYWSGVYSLAVHVAETEGSSC--T 1099
Db 333 ESEINFQAKRKLEGRDIPFGPKELFQFLVETSF--WEKYKQKTNKDVMALGDCLFDT 390
QY 1100 SLELYQVLSAM-----RMLLIATTHAIDMLTDSVVRQLF 1137
Db 391 QQTQTEIGDIWDFSEKTSKTTAEMLNVMORANSWMLTSQLSVHALISVL--ILY 448
QY 1138 LDVL-----DQTKALLVPASVNCIRLGSW---KCTLLILLRQMKRELGS 1180
Db 449 EDNISVHLNATPYVQSLAESAAAEKIPSRVTLSIDKVCRKCTTVDSIASIW----- 502
QY 1181 VDEILGPLEIIEGVQADQQLME---KTKAYVSAFTVYLQKE---MKVSDIFQYQ 1233
Db 503 -----DAPKIVFDITLAQADILSRLLKSAKNLSVLCALVLRNVPGLIKTIGSLRHSN 556
QY 1234 LVINVCETLQEEVIALF-----DQTRHSLALGATEDKDSMETDDCSRRHRDQDVCVL 1289
Db 557 AILKKTINLLEVLVLVGVGSDNSNSGGMVTLAKDAEATSDAT-----I 603
QY 1290 GL-----HLAKELCEVEDGDSMLQVTRRLPIPLTLTLLEVSRLM--KONIL 1335
Db 604 GLPLPLCNFNGNPEVLTLCITVDLIRNFIPEFTWFPILIQSLRLQHVILIQDQKSTT 663
QY 1336 FTEATLHLLTLARTQGTATVAGAGITQISICLPLSVQSLSTNGIAQTSPSARSKLDAP 1395
Db 664 SVSAILKFPILTIAQVHGAQMLNLSGFSTLRALMEF---PDGMSTVLSNDEK-- 715
QY 1396 SWPGVYRLMSLMEOILKT-----LRYNFLPE-- 1422
Db 716 -----SLEKTEKQHWIGIGLAVNTAMHSLGSVSAGADIVESVITGFLEKGY 765
QY 1423 -----ALDPVGHQERTLQCLNAVTVQSLACLEADHTVGFILQLSNFMKEMHFLP 1475
Db 766 MISYVLAADPFS--DDRDKVRLRSQRTWTSIAYLRVETHTLLICALASHRSMVKIM 823
QY 1476 QL-----WRDIQVN--IGLQCACTSLHSRMLQH-----YLQNRNG-- 1511
Db 824 DNDSPURENTIHULAFISGAQRLRSOSHILLCPVAKKEPDSCKRPSFINTKGWF 883
QY 1512 -----DGLPSAVVQVRPPSAASAPSSSKOPADTEASEQALHTVQYGLKILSK 1564
Db 884 SLAPLCVCKPRTIVASIS---TALVNGDTEHNGSVQSPFSVAQIRIVASILK 940
QY 1565 TLAA-----LNHF-----TPDVQCILLDQSLDL--AEYNFLPALSFTTPTF 1603
Db 941 FLCLQAGVVTAEAEVGYVDIAHFPLEPEFIHIGIQDQATVAV---LCDNYSKSEI 996
QY 1604 DEVAAPSFGTLATVNVALNM-----LGEILD---KKKEPLTQAVGLSTQAE- 1646
Db 997 PDEVKKLCMLQOTTEMSLYLELVCAVQCRHHPVGRVDNFSKDKLVKVAEVAHTYLER 1056
QY 1647 GRTTLKSLMF 1657
Db 1057 SIDSILKTIKIAF 1067

RESULT 8
H96803
unknown protein TSM16.5 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C/Accession: H96803
R/theologis, A.; Becker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso,

```

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 angen, N.F.; Hughes, B.; Hultzer, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maitl, R.; Marzall, H.  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Shinn, P.; Davis, J.C.; Davis, R.W.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Venter, J.C.; Venter, R.W.; Tallon,  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:1130712  
 A:Accession: H96803  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2110 <STO>  
 A:Cross-references: UNIPROT:Q9CAQ9; GB:A8005173; NID:96382494; PIDN:AA07780.1; GSPDB:GX  
 C:GeneticS:  
 A:Gene: TSM16.5  
 A:Map position: 1

Query Match 1.9%; Score 175; DB 2; Length 2110;

Best Local Similarity 19.7%; Pred. No. 0.0058; Matches 331; Conservative 233; Mismatches 592; Indels 524; Gaps 83;

29 ALRELQIBALNKHRLLEGLSYKPPSPSAEKYAKNDVASPKELGL----RISK 84  
 169 ALRNLGVD---DGYMLTLESGVDIVVLSLSDNRSGNMAASLARLVLSFCDSTIQ 225  
 85 FLGLDEBOS-VOLIQ-----CYLOEDYRGTRDSVKTVLQDEROSQALILKT-- 129  
 226 ILNSGVKSLIQLLEQKNDIVNRSAAADALEALSNDDEAKCYKAGVHALIEAIVAP 285  
 130 -----ADYVEETCLRCVL-----HLLTF-QDERHPYVE-YACVDLLEKELV 174  
 286 SKECMOGHGOSLOEHANALANFGGRHLIYLGVSOSPLTEPIGDVIGALAYALM 345  
 175 SKYRQFEELYKTEAPTETHGNLATEROVSRMFVQCLRE-----OSMLLEIFLYY-- 226  
 346 -----IFKOP-----ESSENTIPDSVIESILVKILKRPDTGLIBERILEMAASYGN 392  
 227 ---AYEMAPSD--LVLVTMFKEQFGSRQTRHLVDETMDFVDRIGYFSAALIVEG 280  
 393 SLSLCYDDDAKAVLIALITM-----ASADVERERILICLSGCHDKVGEIWEAIGREG 446  
 281 MD-----IESLH--KCALDDR-----ELHQFADG----- 304  
 447 IQLFISFLGSSSEQHOEYAEMLKILTAQVDDSKMAVTAAGIPLVLQLETSGQAKED 506  
 305 -----LICOD---MCLMLTFGDIPIHAPVLLAMALLRHTLNPEETS--VVRKIGTA 353  
 507 AACLMNLCHSEIRCOVERA--GGIP-----AFMLMLKTGGSPSOETSATKLVKLVHTA 560  
 354 IQLNFVYLTRLQLSLASGNDCTTSTACMCVGLSLFVLTSLATL--GNQODIIDTA 411  
 561 DPATINQLMALIL-----GDDPTSKIQVIEVLG-----HVLKSKASQEDILVHRG 603  
 412 C-----EVLADPSLPPELFMGTEPISGIGIILDSVCGMPH 446  
 604 CAANKGLRSVESITSSRETKHTASVLAD-----LFSSRQ-----DICG--H 645  
 447 L-----LSPILQLRALVSGSTAKKYVSFLDKMSFVNELYKHPHDVISHEDTLMRRQ 501  
 646 LAYDDIINPMIKLITN--NTONAKQVAPALDALSRVKNKNNKKSYIAEGD----- 696  
 502 TPKLVPYLGQTNLRIFQTV--GQVMD--DRAVYVREYSSVSWTLTCE----- 549  
 697 -IKSLIKLAKKSSIESAENAVSALANLSPDIAAALAEADVSAFRIIADGSPBGRN 755  
 550 -----IEMLLHVSTADVIQHQRK-PIIDVYHVISTDLSIAD----- 588  
 756 ASRALHQLKNFPCVDVKSQAQRFAILSLVDSIKSIDVSDAFNILEVALLAKTKS 815  
 589 -----CLPITSRIYMLQLRLT-----TVISPP 611

816 GVNSYSPWIALAEVPSLETIVQCLAEHGLTVODKAEIVLSRLCSDQGFILSELIVSRP 875  
 612 VDI-----ASCV-----NCLTVLAAKRNPAKWTD-LRHTGFPPFAHPV----- 650  
 876 KSMVLADRIVNASLSLEVRVSTALLCAKKEKQILTEITDQSGFKLHLVDMIKH 935  
 651 SLSQMSIAGNNAGGNNLNMSEOPQGY-----GVITA--FLRLITTL-VKQ 698  
 936 NSTSFSLETETVPKGF--LEKNVFQDTSFYPPDPKILIGVVALMLCLITVSDAKSK 993  
 699 LGSTOSQGLVPCWFLKEMLPYHKWRVNSHGVE--OIGCLIELIHAILNLCHETD 755  
 994 VIVWAGGLE-----VLVGLKA-----RYNNSAQBEPEDREGIWSALLALM--FQDD 1040  
 756 LHSHTPSLOFLCISLAYBAGCTVINIGIVDTIDVMAAO-----PSDGA-- 806  
 1041 NVSFSSTYMRIL-----PTLAVILG-SDELIDRFAAHAMASLVCTNRGINTLT 1088  
 807 -GGQGGQLIKTVKLAFSVTNNVIRLKPSPNVVSPLQALSOH-----GAHGNNL 855  
 1089 IANSGAVSGIINLIGYSEIILNDVLANEPISVKEPDQVILQHLFEIEDVRIGSTARKS 1148  
 856 IAVLAKYIYKHD-PALPRLAIOQLKRLATVAPMSVYACGNDAARDAFLTRLOSKIE 914  
 1149 IPIVLDRPIPRPGAPQFAVOILIRIADGSDTN--KLMAEAGAVE--ALTKYLSLP 1204  
 915 DNRKIVILEPLVAVETQGLIELPLNLEVKQSDGSKSPSGMMSCLHAYELDSQ 974  
 1205 QDSTEVAISELNV-----LFSNHEL--QNEMALSSQLTAVLRL--GSR 1247  
 975 QDRYWCPLHRAALAFILALW--QDRDSAMLVTRKPFWNILSPFGTSPSETS 1032  
 1248 SARI-----SAGALNELFDENIRNSELIC-----QAVOPLMDIGSVSESE 1290  
 1033 EPSIL-----ETCALIMKICLEIYVYVKSGLSDLSKDTLK-----KFSI 1072  
 1291 QEVALSALILGSSGNTSNTALLIDVBSLLENVILKISSATASBELKINARLCSVFSN 1350  
 1073 E--KRPAYMSGYKSLAVHVAETBGSCTSLLEYQMLVSAKMLLIATTHADIMHTDS 1130  
 1351 KNIRTSASAGCKMPL-ITLMQSERSAV-----EAAVFAIKILL--DDEQHLELA 1398  
 1131 VVR--RQLPLVDVDTGKALLLVASVNC--RLGSMKCTLLIILRWKRELGSVDDELGP 1187  
 1399 AANHIOELLVGLVSG--KNVYILERSLSALKLK-----DRVPRK 1437  
 1188 LTELLEGVLQADQOLMEKTKAKVFAFIVLQW--KEMKVSDFPOXSOLVANCETLQEE 1245  
 1438 LDMVEAGIIRCELELPGASSSICSAVELFRILITNSGVIAARRPDVAKYVEPLFAVLRS 1497  
 1246 VIALPDQTRHSLALGSATEDKDSMETDDCSRSRRDQDGVGLHLAKELCEVDEGD 1305  
 1498 DLTLMGQHSALQALVNLKQOTLEA----- 1523  
 1306 SWLOVRRLPILPTLITLLEVSIRMKNT-----HTEATLHLLTLARTQCATVAG 1359  
 1524 -FSFTPEAIVP-LISFLESSQALQOLGAELLSH-----LTMEDFOODIT-- 1568  
 1360 AGITOSICPLLSVYQSTNGTAQTPSASRSKD--APSWP-----GYRSLMSLMQ 1410  
 1569 ---TQSANVPLV--RLAGIGIISLOETAKALEKISASNPXAVLDAEGIFELSKYLQE 1622

RESULT 9  
 T38774  
 myosin-3 heavy chain, type II, myo2 [similarity] - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C:Accession: T38774; T43276; T43553  
 R:Stelton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
 submitted to the EMBL Data Library, August 1997  
 A:Reference number: Z21751  
 A:Accession: T38774

A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2104 <SKE>  
 A:Cross-references: UNIPROT:O1157; EMBL:Z98762; PDB: CAB11475.1; GSPDB: GNO0066; SPDB: SH  
 A:Experimental source: strain 9721(-); cosmid c48  
 A:Motif: F.; Nakano, K.; Kiyama, C.; Yamamoto, M.; Mabuchi, I.  
 FEBS Lett. 420, 161-166, 1997  
 A>Title: Identification of Myo3, a second type-II myosin heavy chain in the fission yeast  
 A:Reference number: Z22385; MIMD:98119399; PMID:9459302  
 A:Accession: 143276  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2104 <MOT>  
 A:Cross-references: EMBL:AB007633; NID:G2828347; PDB: BAA24579.1; PID:G2828348  
 R:Bezanilla, M.; Forsburg, S.L.; Pollard, T.D.  
 Mol. Biol. Cell 8, 2893-2705, 1997  
 A>Title: Identification of a second myosin-II in S. pombe: Myo2p is conditionally required  
 A:Reference number: Z22572  
 A:Accession: 143553  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1132; 'G', 1194-1303; 'G', 1305-1343; 'K', 1345-1419; 'D', 1421-2104 <BEZ>  
 A:Cross-references: EMBL:AF029788; PDB: AAC04615.1  
 C:Gene: SPACA8.05c; myp2  
 A:Map position: 1  
 C:Function:  
 A:Description: may be involved in cytokinesis and stabilization of F-actin cables  
 C:Superfamily: myosin heavy chain; myosin motor domain homology  
 C:Keywords: actin binding; ATP  
 F:92-755/Domain: myosin motor domain homology <MMO>

Query Match 1.9%; Score 171.5; DB 2; Length 2104;  
 Best Local Similarity 16.3%; Pred. No. 0.01; Mismatches 627; Indels 739; Gaps 89;  
 Matches 323; Conservative 298; Mismatches 627; Indels 739; Gaps 89;  
 100 YLQEDYKGRFDS--VKTIVQDEROSQALIKIADYYEERTCILRCVHLTYFODERH 156  
 3 YLQGN--GSDNNNIKKLVDAEKHNA--VDAEFDERTWI-----WIPDSKE 47  
 157 PYRVEYADCVKLEKELVKYRQOF-----ELYTEAPTEWTHGNLMTTERQVS 205  
 48 SPYKAW-----IYEDLGEKRYKVLGERDGERIVDGFDAKVNPPKDMVDMAA----- 96  
 206 RWFVQCLREQSM-----LEIIFLYAYFEMAPDLDVL-----TKMFK- 244  
 97 --LTLNPSVANNLTQRYEKDLITYSGFLVAVNPPCHLPYGDVVRKQSKQFKE 153  
 245 -----BOFGSRQTNHVLVDETMDPFVDRIGYSAL 275  
 154 TKPHIFGTADAAYRSLIERINGSLVLTGSGAGKETTKVQ-----YLTSV 202  
 276 ILVEGNDIESLHAKLDRRELHOFADGLICODMC-----LMLTRGDIPIHAPVILAW 330  
 203 TDASTSSQOLEKKILETNVLEAFNAQTVRNNSRPFKFIRESNNGSIVGANLDM 262  
 331 ALAR--HTLNPETS-----SVVRKIGGTAIO-----LNVFOYLTLLQSLASGN 374  
 263 YLEKSVIPISSNERYAHVYQLLRGADSLSEFLDXYVDHYSLKNGKLKHI--NGV 320  
 375 DCTTSTACMCVYGLSLFVLTSLHLTL-----GNODIIPDAACEVLAD----- 417  
 321 DDGEKPKCLC-FGLRTLGFNNNEIHSFLIILASIHGN-----IEVASRSGGARF 371  
 418 PSLPQLFVWGTERTPSGLGIILDSVCGMPHLS--PLQLLALVSGSTAKKVSFLDK-- 474  
 372 PSLTQ-----IDOLC-----HLLEIPVDFVNAALHPXSKAGREVIIVARTR 413  
 475 -----MSFYNELYKHKPHDVISHEDTLMBRQTPPLVPLGGQTLRLPQGVGVVM 527  
 414 EGVVHTLQSLAKGLIYENRPHLVKRLNQITWYSGSEH-----DGTIG--VL 457  
 528 DDBAYLVREWYSYSSWT-----LFTCEIEMLHVSTADVIQHCORV 569

DB 458 DINGPEI---PTFNSPEQCLNFTNEKLOQFPHHNVFVLEGEYTERIERIEMWDIDVGNL 514  
 QY 570 KPIILVHKVISTDLI-----ADCLLP-----TSRIYMLQRLTTVISP----- 611  
 DB 515 QPITDAIEK--SRPIGIFSCLEDCEVMPATDATEREKLHLLEFKGKSDIRPKPFSESGF 572  
 QY 612 -----VDVIASCVNCLITVLAARNPAKWTDLRHNGFLFVNAHPS 651  
 DB 573 VLKHVAGDVEYDTKMLEKKRQDPLNACLAAMFKSTN-----S 610  
 QY 652 SLSCMTASAGMNAAGYGNILMNSQOPQGRY-GVTIAFLRLITLVKQGLSTQSGLVPC 710  
 DB 611 HVSLLEDVDSNNSAGSDNI-----EKKGIFRTYSQHRRQLSLM--QLEATQPH-FVRC 663  
 QY 711 VNFVLKEMLPSTYKWRYN--SHCVREQIGCL-ILELIH-----AIIMLCHETDLSHSTPS 763  
 DB 664 I-----IPNNIKOPHNLDKSLVHLQRLCNGVLEIRIAQTGFPPKLFYTERFARYGIL 716  
 QY 764 LQFLICSLAYTEAGQTVINIMIGIVDTIDMWAAQPRSDGAGGQGGQLITVTKLAFS 823  
 DB 717 SGL--KRGVYAKKATIT-----INELKL----- 740  
 QY 824 VTNNAVIRLKP-----PSNVVSPLEQALSOHGAHGNLILAVL-----AKITY- 864  
 DB 741 -PSTVRLGETKVFVKASVLGSLD-----RRNALLVIFNVSFARIRGFLTRRLVYR 792  
 QY 865 --KHDPALPRALQILKRLATVAP-----MSVYACIG--NDAAIRDAFLRLQS 911  
 DB 793 FNRQDAAL--LQHNLRQLKILKPPHWNVFLHLRPLGTOTDYLRLKDLINLON 850  
 QY 912 KIEDMRKIMILEFLVAVETOPGLLEFLNLEVKQSGSKFSLGMSCLHVAVELID 971  
 DB 851 QLESTK--EVANELLITKE-----RVQLQTN 874  
 QY 972 SQQODRYWCPPLHRAAIFLHALMODRDSAMLVLRTPKFEWNTSLPLFGLTSPSET 1031  
 DB 875 DLQEQ-----ALAEKQDILVERANSRVEVHERLSLENQV-----TIADEK-- 917  
 QY 1032 SEPSILETCALMKITCILEIYVVKSLDQSLKDTLKRS-----IEKRAVNSGYV 1083  
 DB 918 -----YEFYAEKQSIIEEDLANKQTEISYSDLSSTLEKKSIIKQD 960  
 QY 1084 KSLAVAHAEETEGSSCTSLLEYQMLVAMRML-----LITATADIMHLLTDSVVRQ 1135  
 DB 961 QITSSKXKLENDYINMADYQHSQHLNLEKALINEKMLNIRELNEKMLRDLDELKQ 1020  
 QY 1136 LFLD-----VLDTKALLVPAVNCRLGSMKCTLLIL--LRQW 1175  
 DB 1021 RSYDTKVOELREBNASLKQOCRTYESQSLASVSKYSETESELNKAELVIFQKEITERYR 1080  
 QY 1176 RELGSV--DEILGPILELLEGLVLODQOMETKAKVNSAPITVQOMKEMKSDIPQVS 1232  
 DB 1081 DQHLKAFQNPETKHNINDVSGPLNSDENIY--STSTTSLKQVDELSLHATKENQUS 1139  
 QY 1233 QLVANVCETLQOEVLIA-----LFDQFHSALGASATEDKDSMET 1271  
 DB 1140 ERKEISEMLEQSIAREEKLRKNSLCOILIAKTKQIDQDEFEIISNA-----DNLD 1194  
 QY 1272 DDCSRHRDQDRGCVLGL-----HLAKELCEVDEDEDSMLQVTRRLPILP 1318  
 DB 1195 KOTNGVLEKNASDFIDFGIKRYEKKISDLNQLQEKERCKVQ----- 1237  
 QY 1319 TLLTTEVSLRKQNHFEATLHLLTLARQOQGTAVAGAGITIOSICLPILSVQOLST 1378  
 DB 1238 -----LKKQTKENSVTQHTL----- 1253  
 QY 1379 NGTAQTPSARSKSLDAPSWGVRL-----SMSLMEQLKTLRYNFPALDPVGHOR 1433  
 DB 1254 DQNSHPSPFBEKHSQGP-----LKRIDGNNDDKIDKDKLKT-----ISKSLDLQGLVBE 1304  
 QY 1434 TLQCL-----NAVTVQ--SLACEADHTVGFILQLSNFMKEWHFHP 1475

Db 1305 ELSNLSYSLSKDLSFTDIGHIIPNSIRKLEKSLSTSELER-----LNASNSDRP-----SP 1356  
Qy 1476 QLMRDIONLGLVLCACSTSLHSRMLQHYONKNGD---GLPSAVARQVRP----- 1525  
Db 1357 DIFKDTQ-----AINMSRKL-----LSNPSNDQSGISSLQKLLNPNESNMEFT 1401  
Qy 1526 -----PSAASAPSSSKQPADTEASEQALHTVQYGLKLKSLSTLALRHF-----TP 1574  
Db 1402 GLKPLSPKISNLBSS--QPSGSPKRS-----GGMELINFDONSSIP 1443  
Qy 1575 DVCQILDQ-----SLDLAEYNFLPALSFTPTPDPSEVAPSPGTLATV-----NVA 1621  
Db 1444 D--PPIVQGRNSVLQTEPEKINLKKKATYKSGILDNKOLSKFSELIQSLKENBELKYL 1501  
Qy 1622 LNMGEIDKKKE--PLTVANGLSGQAE-----GTRILKSLMTMENCPIYLISQ 1669  
Db 1502 TSNLSGSDKMLDFAFLLEDVDPNTTNOIKGFVEKAISSKRAIAKLYSASEELF----- 1555  
Qy 1670 AMRYLRDPAVHPRDKORKEISETLSTLSLSRYPFGAPSPSPATG-----VLPSQP 1723  
Db 1556 -----STEKALREITTKERDLHLGLQ---GPSVPTSPKAPLASQLIIPVD 1599  
Qy 1724 GKSTSL 1730  
Db 1600 GSITNYS 1606

## RESULT 10

B96671  
similar to translational activator [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: B96671  
R:Thellogis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hulzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzbey, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141, MUID:21016719, PMID:11130712  
A:Accession: B96671  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-2698 <STOP>  
A:Cross-references: UNIPROT:Q9XIR5; GB:AEO05173; NID:g5042415; PIDN:AAD38254.1; GSPDB:GN  
A:Gene: F13011.10  
A:Map position: 1

Query Match 1.9%; Score 169; DB 2; Length 2698;  
Best Local Similarity 18.8%; Pred. No. 0.023;  
Matches 385; Conservative 269; Mismatches 674; Indels 720; Gaps 97;

Qy 158 YREYADCVDKLEKELVKYRQOFELKYTEAPMET--HGNLMTERQVSRWFVQGLRQS 216  
Db 844 FKMHLQDLPRLSHMDLS-----ETDIKIPHTPEGMILSRQGV--YVAQITGAK- 890  
Qy 217 MLLEIFLYIYAFEMABDVLVTMFKEQGFGSGQT---NRHLVDETMPFVDRIQYF 272  
Db 891 -----YTKQEPSSNSHLKK-----GLASRETANSGRDPAKLTAKKADKKTAK 934  
Qy 273 SA---LILVEGDIESLHKCALDRRELHOFADGL-----ICQDMDCMLTFGDIHPAP 325  
Db 935 EARELMLKEBASSTRNVRIQKSLSLVHALGEMGLANPVCHSQSLPFLATFLDPLAS 994  
Qy 326 VLLANA-----LIRHTLNP-----EETSSVVRKIGGTALQNLNVQY-----LTR 364  
Db 995 IVSAAAFENLVLAARCTQVPLCNMMLAISTALRLAIDEVDTN--PDFRPSVDKAGKTYEG 1053

Qy 365 LLQSLASGNDCTTSTACMCYVGLL-----SFVLTSELTGNOQDIIDTACEVLADPS 419  
Db 1054 LPERIVNG-----LSSICKSGPLPYDTFFIFIPVL--YHVLG-----VVAAYASVGR 1100  
Qy 420 LPELFMGTEPTSGIIGLIDSYCGMFPHLLSPLOLRAL-----VSGKSTAKKYVSFLDK 474  
Db 1101 LNELCLGQAD-----VANALGYVSKDVHRLACLAINVKCI PAVSKCSLPQNV-----K 1151  
Qy 475 MSFNELYKHKPHVDVISHEDDTLMRQTPKLLYPLG-----QNTLRIPQGTGVGM 526  
Db 1152 IATNIMIALHDPEKSVASADLMARYGHDIGTDSGLFKALSHININVLAAABA----- 1207  
Qy 527 LDDRAVLRWEYSYSWTLFTCEIE-----MLHVSADYVQ----- 564  
Db 1208 LADALHESPSSIQLSTLPSFLYTRDATSGEDVDAGMIGROGIALALQSAADVLTMDL 1267  
Qy 565 -----HCRVXPIIDLHKVISTDLSIAD-----CLPPI----- 593  
Db 1268 PAVMTFLISRALADPNDVGRKMINAGIMIIDKHGKENVSLPPIFENYLNKEASDEBEY 1327  
Qy 594 -----TSRIYMLQRLTYISFPVIVASCVN--CLT----- 623  
Db 1328 DLVREGVVIPTGALAKHLARDPPKVNVEKLEVLNTPSSVQRAVASTCLSPVLRSOL 1387  
Qy 624 -----VAAANPAKVTDLRHTG--- 641  
Db 1388 MKSDKYGERRGAGLACVVMGFGISLKKYGLVTLQALIDRNSAK-----REGALL 1442  
Qy 642 -----PLPFAVHPVS-----SLSQWISAEGNAGYGNLNMNSQOPQGEYVTI 685  
Db 1443 AFECICERLGLFPFPYVYKMLPLLVSPDQVAVREAAFCABAMMSQ---LSAYGVKL 1499  
Qy 686 AFLKLITTVNG-----QLGSTQSG-----LVQCMVVLKEMPSYIKM 725  
Db 1500 -----VLSLKLGLBDKAMRTQSSVQLGAFAPCAPOQLSCLPRVVKLTLEVLTDHP- 1554  
Qy 726 RYNSHG--VREQIG-----CLILELI-----HAIALNCHETDLSSHTPS 763  
Db 1555 KYQAGAGLALQOVSVIKNPEISSLVPLTLALDNPENYTHADTLQTFVNSVDAPS 1614  
Qy 764 LQFLC-----ICSLAYTE-----AGQTVINIMIGVDTI 792  
Db 1615 LALLVPIVHRGLRERSSETKKKAQIVGNMCSLV--TEPKDWIPIYGLLPEVKVLYDPI 1673  
Qy 793 DMVMAAPRSDGAGQGGQ-----LKITKYKAPSYNNVYIKLRPSNVVSPLEQA 844  
Db 1674 PEYRSVABARAVALSIRGEGEDNFPDLVPMWLPETLK---SDTSNERY---GAAQGLSEV 1726  
Qy 845 LSGHGA--HGNLILVLAKEYIYHKHDA-----LPR-LAIQLKRLATVAP----- 887  
Db 1727 IALGTDFENILPDLRHCSHOKASVRDGYLTLPKFLPRSLGAQFOKYLQVLPAIIDG 1786  
Qy 888 -----MSVYACIG--NDAAIRDA-----FLTRL 909  
Db 1787 EPIFGRLTJSELPFDSLTKLVLYVDVLGLADENSVADALGAGHVLVEHHAATTSPL 1846  
Qy 910 QSKIED-----MKIKWILEFL-----TVAVEFQPGILEFLNLYKQSGD--GSKFSS 956  
Db 1847 LPAVEDGIFNDNMWIRQSSVELLDLLEKVVAGTSGKALLE-----GSGDDRCASGTEA 1898  
Qy 957 LGMMSCUAVVELIDSQOODRWYCPPLHRAALFIALHMDRRDSAMLVYLRTPKPFEN 1016  
Db 1899 QG-----RAITDILGMDRNE-----VLAALVNRVDSLSVQALAHVKT 1940  
Qy 1017 -----LTSPLFGTL--SPSESTSEPSIETCALINK-----IICLEIYVVK 1056  
Db 1941 IVANTPKLKEIMPLIMSTLSSLASPSSERQYAGSGLVYKGERVLPFLIPILSK 2000  
Qy 1057 GSLDQ-----SLKDTLKFSIEKRPAYMSGYKSL-----AVHAYLEGSSCT 1099  
Db 2001 GLKDPDVYDKROGVCIGINEVASAGRSQLSFMQLIPTIRFALCDSALVEEBSAGLAFS 2060  
Qy 1100 SLLEYQMLVASMRMLLIATTHADIMHLDTSVVRQLFLVDLDTKALL-----L 1149



Db 1063 LPOSESISQNVRL-----AEPSS--NCDS--ILKINAPGPOTTTSHL 1106  
Qy 1037 LE--TCALIMKICLEITYVVKSGIDSLKDTLKFFSIEKPAVSGVYKSLAVVAATE 1094  
Db 1107 IEKTPAAVAVAN-----VKAAVESVIS--KSSSLAKPABEKAITITMLAGLVSKD 1154  
Qy 1095 GSS--CTS-----LLE-----YQMLVSAWML 1114  
Db 1155 VSSISCTSETGFDFORIFRONANISLGAVSEITLQRLREPIENOVQGFSTANSOKA 1214  
Qy 1115 LIAITTHADIMHLDTSVVRQFLFDVLDGTAKLLVPASVNCRLSGMKCTLLILLRQW 1174  
Db 1215 SFIAKQKLDVTWDTMKVIAIQMISQITDGD---FSSASISSETPKINIDKAREVV--- 1266  
Qy 1175 KREIGSVDEILGPLTEILEGVLOADQOLMEKTKAVSAFITYVLOMKMKVSDI----- 1228  
Db 1267 AAEFGINAVSOKALEVANKWMSDISLPEKIQWNI--SSNVRVLADSSYNCDSIIQKINA 1325  
Qy 1229 --POVSOLVNVCEITLGEVIALFPDQTHSLALGSATDKDSMETDDCSRSRHRDQGV 1286  
Db 1236 PEPQSAVYDQOISEHQTEVLANKISAVSV-----ISNDVLSKRSEDEKAVI 1374  
Qy 1287 C-VLGLHLAKELCEVDEGDSW-----LQY--TRRL--PI----- 1316  
Db 1375 SNIANLITISCLASWSTNSPEFORDILBPONANVLGTPTASQVFTNNLQPIQSEVQG 1434  
Qy 1317 -LPLITLTLEVSLRKQNLHF-----TE--ATHLLITLTARTOOGATAVAGAIT 1363  
Db 1435 FWSSTNSQEKASLVIRQKLVKYDAMKMAETIAYOSLNSNTSETYLESCSPKDSIR 1494  
Qy 1364 OSI-----CLPLSVYQLST-----NGAOTPSASRSKSLAPSPGYR 1402  
Db 1495 ESIMASFEVSEBHVOKTEILSKSELSTIVLPAVEYETVO---NIALSEPNN--- 1546  
Qy 1403 LMSMLEQLLKTLRNPALDFVGVHGOERTLOCLNAVTV-----OSLACLEBAD 1454  
Db 1547 ---CESITGLKLNAPDQSAVMDIMNEQRKLTIVSNIGSABSVISNDSLEKLEVD 1601  
Qy 1455 HTVGFITLQLSNPKEMWHHLPOLMRDIOVNIQYLCQACT--SL--HSRKLQHYLONK 1509  
Db 1602 KSV--ILKISREI-----VSHDLTALSSCSDFNLQESNTQEKADIFLKT 1646  
Qy 1510 NGDGLPSAVNAORVORPSAA--SAAPSSKOPADTEASEQOALH-----TVQ 1555  
Db 1647 NSQMLIEREPIEQOVGFSTANSSEKQEMFLEKYE--TIHMLQTFPSASVSEITQ 1704  
Qy 1556 YGLKILSKTLAALRHFTPDVCOIL-----LDQS--LDLAEY-----N 1591  
Db 1705 RQFPAIV--QSLAALRSIQLTRREILCAFGISNEQIBQTFRGLDVEDMSHIDIPCALDN 1763  
Qy 1592 FLFALSTFTPTFDSVAPSEFTLLA-----TVNALMLGELDKKKEPL 1635  
Db 1764 LLVNLRIY---NAAPNVFGLILSRPDNOETSTVEEKRIQPLNLQSLDQTV--M 1818  
Qy 1636 TQAVGLSQAGSTRRLKSLMFTMENCYLLISQMRYLDPAYH-----PRD--- 1683  
Db 1819 TSSLRSRDERIEVKVSNITISLISEN--LGDLSQAVQ--LAOPNWSDETEKFEIDPRNLL 1876  
Qy 1684 -----KORMKQELSELSTLSSLSRYFRGAPSSPATGVLSPQOKS 1726  
Db 1877 NRNVPESEBSIQSFWKTSQLAEBASSTISEKLSMLQSEF-----VVSAAKQVS 1925  
Qy 1727 TSLT 1730  
Db 1926 TSLT 1929

## RESULT 12

T28915  
hypothetical protein C13F10.4 - Caenorhabditis elegans  
C13F10.4:29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C13F10.4:29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C13F10.4:29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004

R:tin, A.; Wohldmann, P.  
submitted to the EMBL Data Library, April 1997  
A:Description: The sequence of C. elegans coemid C13F10.  
A:Reference number: Z20543  
A:Accession: T28915  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2076 <TIN>  
A:Cross-references: UNIPROT:O01485, EMBL:U97006, PDB:AA047965.1, GSPDB:GN00023, CESP:C1  
A:Experimental source: strain Bristol N2; clone C13F10  
A:Gene: CESP:C13F10.4  
A:Map position: 5  
A:introns: 114/2; 264/1; 319/3; 358/2; 664/2; 756/3; 840/3; 1022/3; 1119/3; 1840/1; 2056

Query Match 1.8%; Score 163.5; DB 2; Length 2076;  
Best local similarity 19.2%; Pred. No. 0.035;  
Matches 274; Conservative 161; Mismatches 459; Indels 533; Gaps 62;

Qy 446 HLLSPILQILALYSGK-STAKKYVSFLDKMSFYNELKHKPHVISHEDGTLMRORPK 504  
Db 156 HMMTLLSMVVALSGDYTHKKHISAKN-----SLQDRILCK--TP-- 196  
Qy 505 LLVPLGGQTNLRIPQGTGQVMDRAVLYWMEYSYSMTLFTGEIE---MLHVVSTA 560  
Db 197 -----NKKI-----ASLECLTALVGFH-----TPYITTELEASCTMCIKILEGS 235  
Qy 561 DVIOHCORVPIIDLHKVISTDLSIADCLLPITSRIYMLQRLTTVISPPVDVIASCVN 620  
Db 236 TYELRCVAVK-----FTAQLLATSKPPGAVIRAKS 267  
Qy 621 CLTFLAANPAKWTDLRHGTFLPVAHPVSSLSQMSABGMNNGVGNLMSF---QP 677  
Db 268 NQTFPV--RPASVDTLN-----LASSFLKGGIGGFLKGSSTST 307  
Qy 678 QGEYGVTFAPRLITTVLKQGLSGTQSGLVPCVWFLKEMPSYHKRVNSHGVREOIG 737  
Db 308 MGRSDIRL-----GVISCYEWMREK-----G 329  
Qy 738 CLIELIHAILNLCHETDLSHSPSLQFLCTG--LAYTEAGQYVINIMIG-----V 789  
Db 330 SAWLEK--HLIIVCGHMDLASK-----CGHLATYQASHVSEALTRRCISFLR 378  
Qy 790 DTIMWMAAOPRSGABEGGQGLIKTVKLAFTVNNVITLKPPSNVSPLEQALSQHG 849  
Db 379 QTIGSL-----GENAQTLACKRLGVLISQYVDLVSIGTDNLDSSVDS--SDYG 426  
Qy 850 AHGNNLIVLAKYIYHKDPALPRLAIDLKRLATVAPMS-----VYACLGADA 898  
Db 427 S-GYAIIVILQD-----ISLVROIGTSMVSLFTEATGIMEHIFKCLTHPL 471  
Qy 899 AAIIDAPLTRLOSKIEMRIKVMILEFTVAETQPGILFELNLEVKGSGDSKEFSLG 958  
Db 472 ASARYA-----AAMCIRCIATAV---PMLMTPLID----- 498  
Qy 959 MMSCHAVLELIDQOQDRVYCPPLHRAALAF---LHALQDRDSAML--VLRTKPKFW 1014  
Db 499 --RCLPRLDQSSS--RAISGFMAISALAASTDSKLDIPIYAKPLKV 544  
Qy 1015 ENLSPFLGTSPSESTSEPSI--LET-----CALI-----MKIICLEIYVVKSL 1059  
Db 545 LDLAEMKRT-----STQPKLTAKLESQNNLILYALHGLSPVMEKHLPRVILKMKAF 599  
Qy 1060 DQSLKDTLKFFSIEKRFAY-----WSGYKSLAVHVAETEGSSCTSLLEYOMLVSAWM 1113  
Db 600 ARSAKEASENSRQDGFQCAMIAGALSVMEAVASQPELSTNNALDMKPIECSL 659  
Qy 1114 LLI-----IATTHADIMHLDTSVVRQFLFDVLDGTAKLLVPASVNCRLSGMKCTLL 1168  
Db 660 VMMSQVGNLISYNGEMQANSVVRIRLY-----RLTLILP----- 695  
Qy 1169 ILLRQMKELASVDEILGPLTEILEGVLOADQOLMEKTKAVSAFITYVLOMKMKVSDI 1228



Db 696 -----HKSFEVSAAAL---LRELVAIDITLSDNSOSTLTLTSLPISQFTGVCEKILISPYDA 747  
Qy 1229 PORSQVLAVNCTEIOEYIALFDQTRSLALGATPEKDSMETDDCSRRRDQDGVCV 1288  
Db 748 TDVSMV-----EDLLQTPIS-----SVSLGNIEEDLSNLIIRASAS-----782  
Qy 1289 LGLHLAKELCEVDEGDSWMLQ-VTRRLPILPTLLTLELVSRMKONLHTEATLHLTL 1347  
Db 783 -----OIGDTWPENDSEPLCTLNTALTITCKVFPVLNNKHKOITHEPMDTI 829  
Qy 1348 ARTQOQATVAVAGITOSICPLLSVYQLS-----TNGTAQTPSARSKSLDAP 1395  
Db 830 QKSR-----NAGRKQAILVNALTAKLISYKTLCEGRGHKLDBETLQ-----BASFDL- 876  
Qy 1396 SWPQVRLSMQLKTLKRLNPLPRLDVGCHQETLQCLNAVRVQSLA-----1448  
Db 877 -----ISSLSNSCPWTRLVG--AEAL-----ARLSQAVNSPPFVASIAQYCPDKL 920  
Qy 1449 ---CLEADHTVGFILQISNFKEMHFLDPLQIMRDIQVNLGYLCOACTSLHSRKLQHYL 1506  
Db 921 NSCQDEINRS-GHYLALGCL---HRHVGSL-----946  
Qy 1507 QNKNKGDLPSAVAGRVORPPSASAPSSKOPADTEASBOAL-HTVOYGLKILISKY 1565  
Db 947 ---GSG-----QHMLTGVSVLALAESKMPKVCATVAMALIAETSGMFRFVET 996  
Qy 1566 LAALRHPFDPVQCILLDSDLAELNPLFALSTFTPTDSEKVASPFG---TLATVNVVA 1621  
Db 997 -----TSSCLKL-----ISPTFVVDVVOGISKCLTALTTCVGP 1033  
Qy 1622 LMLGELDKKEPTQAVGLS-----TOAEGRTLLKSLMFTMENCYLLISQAMRYL 1674  
Db 1034 LSCGVIDGVATSLAACAIQSHSDPVOQAEIATISGLQMHF-----APRY- 1080  
Qy 1675 RDPVHPDRQKRMKQELSELSTLISLSRYFRGAPS-----SPATGVLP 1720  
Db 1081 ---VH-----MAQLVVDISSLSSTHVLVIRROVSCLRLQVORSEKVRNHAQVLP 1129  
Qy 1721 SPQKSTSLSKAPES-----QEBLLDQV 1745  
Db 1130 QGIYDVTNKKKFLPESGLEALFGMLDTEVNEKELRCHLQETLISLVQ 1176

RESULT 13  
AS7036  
tailin - glime mold (Dictyostelium discoideum)  
C/Species: Dictyostelium discoideum  
C/Date: 03-Oct-1995 #sequence\_revision 03-Oct-1995 #text\_change 29-Oct-1999  
C/Accession: AS7036  
R/Kreimeier, M.; Gerisch, G.; Heizer, C.; Mueller-Taubenberger, A.  
J. Cell Biol. 129, 179-188, 1995  
A/Title: A tailin homologue of Dictyostelium rapidly assembles at the leading edge of cel  
A/Reference number: AS7036; MUID:95213284; PMID:7698984  
A/Accession: AS7036  
A/Status: Preliminary  
A/Molecule type: DNA  
A/Residues: 1-2491 <KRE>  
A/Cross-references: GB:U14576; NID:G797298; PIDN:AA46586.1; PID:G797299  
C/Genetics:  
A/Introns: 131/1; 222/2

Query Match 1.7%; Score 155; DB 2; Length 2491;  
Best Local Similarity 17.9%; Pred. No. 0.19;  
Matches 276; Conservative 227; Mismatches 501; Indels 542; Gaps 72;

Qy 560 ADVTQHCQKVRPIIDLVHVKVISTDISIAD-----L 590  
Db 408 ANQOPQOI-PIYDLKSGALRATDILLIGELGFRSGTATPONTFRSPTTLTPQOFKHQL 466  
Qy 591 LPIIRITM-----LQRLTIVISPPVDVYASCVNGLT-----VLAARPAKVMYDILRTG 641  
Db 467 ISHTNAMAIAAQGLFQDMWT--PPPTGIGIAAFQQAIRRAQIIMAEIN--TVGTAAKNAG 522

Qy 642 FLPEVA---HPVSLSQMISABGNMAGYNLMMNS---EQPOGEYGVITAFLELITTLV 695  
Db 523 YFPDMSFSEIITGVAFKSLSESMARLLAISTIGTDCDEKSKQAAATELFNNQSVLTIM 582  
Qy 696 KG-----QLGSGTSGQIVPC-----VMFLKEMLPYHKRMYSNGV 732  
Db 583 MAACDNEYVDSSSKLIECAKXVSAIADMLVGNKSEFIDELLGOITLTKSTSL 642  
Qy 733 REQICILILHAILNL-----CHE-----JDLHSHTPSLQFL 767  
Db 643 TSD-----BLSTTEMASTSCHPSRKOITNTOSALNQSNAALTFPSGELPEQDY- 695  
Qy 768 CICSAYTEAGQTV--ININGIGVDITDMMAQPRSDGAEQOGOLIKT-----VK 819  
Db 696 ---NLNARSDIIESVNLINLVAMDSEERYKISTNSNGE-VGEGSILAGTNLTFEPAT 751  
Qy 820 LAFSVTNVIRLKEPSSNVSP--LEQALSQGHAGNNLVAVLAKYIYHKNDPALRLAIQ 877  
Db 752 VANDLTNAIMTMR--SNLKMPDTVMESYKQVAGHANRLI-TCTKAVASRAD-----TQ 801  
Qy 878 LKREL--ATVAPMSVACIGNDAAAIADAFILTRLOSKIEMDRIKVM--LEPLVAVETQ 933  
Db 802 SQQLFENSTNAVSESVNLNHCN---SYIKNDEQAHPTQIVETAGHLOFTLQNNSTD 857  
Qy 934 PGLIELEFLNEVKGSDGSEKFSIGMWSCLHAYLELIDSOQOORWCP-----981  
Db 858 AGKACTISLR-----DYSKEM-----IAQVSSLSTBRTSQVLPDANGITLLGAK 905  
Qy 982 ---PLHRAAIAFLHLMODRDSAMLVTRTKPFW--NLTP--LEFGLSPSET 1031  
Db 906 DVSDALSKLWVGIKVILDKSEATQWELLTLOKOSLPPMNLVSTCKRPAPKISDNQK 965  
Qy 1032 ---SEPSILETCALIMKI-----IC-----LEIY-----YVVKSLD 1060  
Db 966 QRLTFSSDAAQSQVKIMKKAQEAIRICGHIETLEALEVDSTIADETTEIATAGFLD 1025  
Qy 1061 Q-----SLKDTLKKFSIEKRFAYWSGYKSLAVHVAETGSSCTSLLEYQ 1105  
Db 1026 AVSGTREGAELLMVAIKD-LNKVNE-----LVTDIRVAPRL-GDLVGSATESA 1075  
Qy 1106 MLVSAMKMLIIRT--HADIMHLTDSVY-----RRLFL-- 1138  
Db 1076 SVAISAKTILICATGQVQTKMGITKQIMDIEQILRASRSVSNPNDRSLLDRR 1135  
Qy 1139 --DVLDGKALLVPAVSNCLRLGSMKCTLLILRLQMKRELGSVDILGLPTE-----1190  
Db 1136 SNDVISITAAVSTANVDCKEDEASADISNL---SLKMSLESILSGPTEEPAYV 1191  
Qy 1191 ---ILEGVLDQDQLMERTKAKVFSAFTVLQMKEMKVSIDIPQYSQ--LVINVCTEQ 1243  
Db 1192 EIASSTKALNAAQVVAARNKRLKGLASAKITASALSTVSHAQNALVLTENEXTK 1251  
Qy 1244 EGVALDQDRHSLALGSAFEDDSMETDDCSRRHDDQDGVCVLG-----HLA 1294  
Db 1252 NALIA-----STVALGG-----QIGLDFSKRIANYDPIYDQNLINQAKSVEDETLV 1300  
Qy 1295 K-----ELCEVEDGDSWMLQVTRRL--PILP-----LTLLEVSLSMKONLHTE 1338  
Db 1301 KVGSLGDDGNNTIC--DEAVDRITTEATRSIDKILPIDTSGLQTNALHEMLHQSL-----1354  
Qy 1339 ATTHLLITLARTOGA-----TAVAGAGITGS-----1365  
Db 1355 ---LATQASKRIGITSMLVNSKNSDVGSGSTDAERLIEMIKAHVHCSISTYN 1410  
Qy 1366 ---TCLPLLSYQLSTNGTAQ-----1383  
Db 1411 PDILIPAKSKILDASQMLTAQADVNVHLSHAATIACTQOLLGITRERASQFNEQDQOV 1470  
Qy 1384 -----TPSASR--KSLDAPSPGVYRLSNLSIMEQLKTLR-----Y 1417  
Db 1471 QVRDGIYSQOQLAHATISSARAVKSVTSKE-PG-----AKAMISQSLKQLESALNNLLIT 1525  
Qy 1418 NFPEALDFGVHQ-ERTLQCLNAVRT-----VQSLACLEADHTVGFILQISNFKEM 1470



```

Db      1526 SSVPSRSGIGIADENKLMSTCRSVSTASQULIISASSCSQKPD-----IELSLISEN 1580
Qy      1471 HFLHPQLMKDI-----OYNGLYCOACTSLIHSKRM 1501
Db      1581 AVLMTNSIKDIIKVTSSMPGVNFCCEAIEIAORAISDLSSVALSVAVGSPSSANNREG 1640
Qy      1502 LQHY-----LQNKDGLPSAVALQVORPPS---AASA-----APG---SSKOPADT- 1543
Db      1641 LSHVESQRLVDVTKITGTGINDLKASRQSPBAGISAKALSPFASLVVTKPKALTA 1700
Qy      1544 -EASEQOALHT-----VOYGLKILSKTLAALRHFTPDVQOILLDQSLDLAEYNFLFALS 1598
Db      1701 PDAAQNDLVTESKNVGSILKLCQASLIASSNPSEKTYQIIVKCVASE----- 1751
Qy      1599 TTPFDSEVAPSFGTLATVNAVNLMLGELDKKEPLTQAVGLSTQAEFTTLKSLMFT 1658
Db      1752 -----AMSLVAVQISSGVNLKELDESILDRIRKSV-VQTSAKDA----- 1789
Qy      1659 MENCEFYLLISQAMRYLRDPAVHPRDKO-RMKQELSESTLSSLS 1703
Db      1790 -----PKQSENRYQOEYKEBELSLNTKLA 1813

```

## RESULT 14

```

probable ubiquitin ligase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T37964
R:Devlin, K.; Churcher, C.M.; Wood, V.; Bartrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z21757
A:Accession: T37964
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3227 <DEV>
A:Cross-references: UNIPROT:O13834; EMBL:Z99531; PIRN:CAB16714.1; GSPDB:GN00066; SPDB:SF
C:Genetic:
A:Experimental source: strain 972h-; cosmid cl9DS
A:Gene: SPDB:SPAC19D5.04
A:Map position: 1

```

```

Query Match      1.7%; Score 154.5; DB 2; Length 3227;
Beet Local Similarity 18.3%; Pred. No. 0.31;
Matches 347; Conservative 255; Mismatches 665; Indels 631; Gaps 89;

Qy      54 YKPPSPSSAEKVK--ANKDVASPLKELGIRISKFLGLDEQSVQQLQCYLQEDYKGRDS 111
Db      12 YSQPPRAAEFRIRQANEVTS---DLGL-----VSLCSFRKNDMPYPRGD 55
Qy      112 VKTVALQDERGQALILKTIADYY-----YEERTC-----ILRCYALLLTYFQDERH 156
Db      56 LYSWVPVLRPDALEIRIVEHYSKDKVQTKPFDSDTLILLELTFSAHLISHCAN-RS 114
Qy      157 PYR-----VEY-----ACVCK-----LEKELVSKYRQOF-----E 182
Db      115 IYNSTVYLEYLINSSVLEVIDSTLALLHLIVQKATISRGKQPLSLODRFLRFLMFLPQ 174
Qy      183 ELVYKTE-APTWETGNLMTERTOV-SRWVQCLREOSMLLEIFLYAYAFEMAPSDLVLT 240
Db      175 DAMTKGFSQNYET--LIFSNEIPEW--CSLEIS-----YKSSPSKDFSSA 217
Qy      241 KMFKEQFGSGSRQTNHVLVDETMDFVDRIGYFSALILVEGMDISLHKCALDRRRELHQ 300
Db      218 SQNPSEGS-----ILKLPYKVL-----GKPIBELLVKTLHDNQIPRQY 257
Qy      301 AODGLICQDMQMLTFBDIPHHNAVLLAMALLRITLNPBETSSVVRKIGTATQANVPO 360
Db      258 SFD-----LLVSLMRQNLVDINRRRLMIRIGLALSNIYVA 294
Qy      361 Y-----LTRLQSLASGNDCTTSTACMCV----- 385

```

```

Db      295 HSOAVQTRFLIADPEITTHLANLVSPPVDLQFNKAVCFEBCFAFFPKSMIPSVIASLN 354
Qy      386 -----YGLLSFVL-----TSLEHTIGNOODIIDTACEVL-----ADPSLBELFWGTEPTG 433
Db      355 VSVSYGLMMNLVRFDSKLENPNFYEREYDSEYDFLQFWTSSPLGGNMACSAGLSL 414
Qy      434 GIILDSVCGMPHLLSPLOLLRALVSGKSTY-----KKVYGFIDKMSFYNEL----- 481
Db      415 GYHLSVTKPQATYYVVASIYVLDHLIDYSNAFAFDPSBSKDLMDLVDRVQYLELAGLQDI 474
Qy      482 -YKKEPDV-----ISHEDGLMRQTPKLLVPL-----GGQTNLRIPQGTVGVMLEDR 530
Db      475 KSGKGNPEIVANMDYALSYDRYFLKVLKLVHLIDGSGSVEIR--NLIDSLISL 531
Qy      531 AYLVRMEKSYSSWTLFCEIML--LHVSTAVIQHCVK-PILDVHKV----- 580
Db      532 AFLLEHHEVVS-NLFASSTNIMSTFIHNEPTCGIIEKRLSHAFAVAVRKLINSSDA 590
Qy      581 STDLSIADCLPIRSRY-MLQRLTVYISPPVVIASCVNCLTVLAARNPAKXTWD--- 636
Db      591 ITSIFLAFGALCINSQGFDFLEK--NPQLPESIFSLNCKSLISSDAAIIGTYIDE 648
Qy      637 -LRHTGLPFVAHPVSSL-----SQMISA--EGMNAAGYGNLNNSEOPQEGYVTI-AFL 688
Db      649 LMRHQ---PSLKQDPIVKMIFKACQVGSALDNFNPQIYNA--KBYVYLLYLETFSSFL 702
Qy      689 RLITLVKQGLGSIQSGLVPCVMFLKEMLPV--HKVRYNHGVRQIGCLILELIH 745
Db      703 ENIIT-NEGHAARYLISKGIYSHVNLIOHPLAFGFISSAFNSFFV-----LH 751
Qy      746 AILMLCHETDHSHTSLOFLCISLAYTEAGQTVINIMIGVDTIDMWMAQPRSDGA 805
Db      752 -----HAYFDABEV--FRPLDCLITRCEEGTEFTIYISKATIS----- 791
Qy      806 EGQGGQQLIKTVLAFSVTNNVRLKPPSNVVSPLQALSOQA--HGNMLIAYLAKY 862
Db      792 -----LIDSMNGHEDANNPLHFSIVGLTLTPALFSSHAALKKAGN--LPVLQ 841
Qy      863 IYKHGDPALPRLAIQLKRLATVA-PMSVYACLG--NDAAIRAPFLRLOSKEIDMRK 919
Db      842 I-----SPSRVAGIDILICNIKSIATSDIHICLGVSDPVLCSLTTIYVDKQEKKEFE 897
Qy      920 VMILE-----FLTVAVETQGLLEFLNLEVEKXGDSQSKESIGMNSCL 963
Db      898 TKKELIYODSFCFPQINRSNFSQIAY---GVSKFTSLTRALGNISVQDFN--EYMI 951
Qy      964 H-----AVLELIDSOQODRWYCPPLHRAAIAFLHALMQDRDSAMLVLRKPKWE 1015
Db      952 HKGSLNALVVDDELINSSKQITSHPSL--SIASLEA-----SLIFVLGASSIRE 1001
Qy      1016 N-----LTSPLFG-----TIS-----PSETSEPSILETCALIMKI 1046
Db      1002 DSKVTLVLLISRLGGCRTVDVLISINETVSGFRLSDRDPKSKNRVLLALSTSLNLT 1061
Qy      1047 ICL-----EIVYVAGSLDQSLKDTLKKPSIEKRAVNSGYKSLAVHAATEGSSCSLT 1102
Db      1062 ILVFTSADPNSEISKTLAMALKSEFDMTF-----NNSGSKLHVL 1102
Qy      1103 EYQMLVSA--WRMLIATTHADIMHLDTSVVRRLQFLDVL-----DGTKALLVPA 1152
Db      1103 HARFISVLMNR-----SADDHL--PYIRALLTVNLNSCYQFEQDITKN---VVD 1149
Qy      1153 SVNCLRL--GSKKCTLLILRQMKRELGVSDEILGPTLELIGVLQADQOLMEKTKA 1208
Db      1150 SINMLRTSIANGDIKEPDLVYTDNTNNSNF-SLEETNASVYDMDES-----EKEN 1199
Qy      1209 KVFSAFIVLQMKMKVSDIPQYSQV-----LWVCETL-QEAVIAL 1249
Db      1200 GIFOAYL--LKMPNDIVSQFEMLKSQIETLVQNASYEGDLMQNLCDPLTYRRDVOQ 1255
Qy      1250 FDQTRHSIALGSAFEDKDSMETDCCSRSRHRQDRGV-----CVLG-LH 1292
Db      1256 NADVQFVTSGLIYEIKKLAGNSTCKAKNQLGPAVGLLSLFIHSDPTQNKAKNCLVSLN 1315

```

QY 1293 LAKEI CEVDESDSWLQVTRRLPILPTLLTLEVS-----RMKONLHFEATLHLLTLAR 1349  
 DB 1316 FFELEHLSLNNG---LPBDSKHTSIVCIIYLLLEVLAAASKRDEFEFNSDCSLK----- 1368  
 QY 1350 TQOGATAVAGAGITQSICLPILSYQVLTSTNGTAQTPSARSKSLDAPSPWGVRLSMSIME 1409  
 DB 1369 -----TDGAIIVDLASQKH-----MSSVI 1388  
 QY 1410 QLLKTRVNFLEPBEALDPVGNHOERTLOCLANVTYQSLACLEADHTVGFILQISNPKKE 1469  
 DB 1389 TLLSLSNAN-----LGV-----VSAFRRVVLTLTSASEMHT-----PVKLSG----- 1426  
 QY 1470 WHFHPOLMRDIOVNLGYLCACTSLH-----SRKK-----LQ 1503  
 DB 1427 ---LPSLPKAMRACSGF-----CNSLHIPTISILRLLPEPEVELMFPDDLVIKFLQ 1478  
 QY 1504 HYLONKNGDGLPSAVAQVQRPSPASASAPSSSKQPAADTEASEQALHTVOYGLKLILS 1563  
 DB 1479 GRARKTEHLHGFIRANAENVLSSPEC-----FIKIL- 1508  
 QY 1564 KTLAARHFTPD-----VCOIILDQSLDAEY 1590  
 DB 1509 KDCCLVGHFTPESEHYVELKESLPGVLQNGQTDLPSEKQMSVYVSFLDELMDLTE- 1567  
 QY 1591 NFLFALSTPTPTPDEVAVSFGTLLATVNALNMLGEL 1628  
 DB 1568 ---TRQSDRSRSPNSEFTPENDSLVMNVVFLQCTEL 1601  
  
 RESULT 15  
 T00637  
 hypochetrical protein H\_GS541B18.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 05-Nov-1999  
 C:Accession: T00637  
 R:Hinds, K.; Sutterer, C.; Becker, M.; Hawkins, M.  
 submitted to the EMBL Data Library, January 1998  
 A:Description: The sequence of H. sapiens BAC clone GS541B18.  
 A:Reference number: Z14196  
 A:Accession: T00637  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Release: 1-1922 <HIN>  
 A:Cross-references: EMBL:AC004013; NID:G2781380; PIDN:AAB96867.1; PID:G2781381  
 C:Genetics:  
 A:Map position: 7621  
 A:Note: Intron positions not resolved (incomplete sequence)  
 A:Note: H\_GS541B18.1  
  
 Query Match 1.7%; Score 151; DB 2; Length 1922;  
 Best Local Similarity 18.7%; Pred. No. 0.23;  
 Matches 304; Conservative 223; Mismatches 561; Indels 534; Gaps 69;

QY 338 NPEETSSVVRKIGGTALQIUNVFOYLTLLQSLASGNDCTTSTACMCVYGLLSFVLTSLE 397  
 DB 544 --SQECCVKNRREIEIOLN--EVIETLQOELANIGK-----TSMN 581  
 QY 398 LHTLGNODIITDACEVLADPSLPFLWGTPEPTSGLIILDSVCGMFPPLHSLPILQILRA 457  
 DB 582 AHSISEADSLKHQDVLVIAEKL-ALGOVETANEEETPKNVKATETNPKNQLTOELFS 640  
 QY 458 LVSGSKTAKKYVSF-----LDKMSFYNELYKHKH-DVISHEDGTLMRQTPKLLYP 508  
 DB 641 LKRESEVEKIQSIPENSVMVAIDHLS-----KDKPELEVVLITEDA-----LKS 684  
 QY 509 LGGGTNRI--PGTQVQWMLDRAYLVKWEYSSWTLPTCELEMLHVVSTADVIQHC 566  
 DB 685 LENQTYKSEENKSGIINLETR-----LLQEST----- 715  
 QY 567 QRVPPIIDLVHKVISTDLSIADCLPI-----TSRIYMLQRLTTVVISPPVDVIA 616  
 DB 716 -----VSAXDLELTQCYKQIKMDQEQGFETMLQKKIVLQKIVE--KVAA 761  
 QY 617 SCVNCITVLAARNPAAKWTDLRHTGFLPVAHPVSSLSQMIASGMAAGYGNLANSQ 676  
 DB 762 ALVSQIQLEAVQEVAKFCOD-----NQTSSPERT-----NIONNOL 800  
 QY 677 PQGEYGTIAFLRLITTLVKGQ-----LGSTQSGVPCVMFVLKEMLPYHKRY 727  
 DB 801 REDLSGDISALTRLISELSQVEHNTSLILEQOVELAKXVLKEXKLLLEQLLEG 860  
 QY 728 NSHGVRE--QIGCLILHAILINLCHETDLSHSTPSLOFLCISLA-----YTEA 777  
 DB 861 NEKKOREKEKRSPODEVELKTTTELPHSNE--ESGFNELEALRAESVATKAELASYEK 919  
 QY 778 GQTVINIMGVDTIDMVAQAQRSDAQEGGQGLIKTVKLAFSVTNNVIRLKPSSNV 837  
 DB 920 AEKL-----OBEELVKB-----TN----- 933  
 QY 838 VSPLEQALSO--HGAHNNLIYALAK-----YIYHKGADLPRLAIOLEKRLATVAPMS 889  
 DB 934 MTSIQKDLQSVNRHIAEAKKSLILEKEDTEVQESKACMFPLPKLSKSLASQVDGT 993  
 QY 890 VYACLGNDAAAI--RDAFTRLQSKIEDMRIRKIMLEFTVAVETOPGLIEFLNLEVK 946  
 DB 994 LKISSNQTPQIIVXKAGIQINQSECSSEVEIISQFTEKLEKQOE----- 1041  
 QY 947 DSGDSKSEFSLGMMSCIHAVLELDSQ-----QDRVYCCPRLHRAATAFLHLM 996  
 DB 1042 -----LHAA-ETLDMESRHSISETTLKREHYVAVOLL--KEEGTLKAVI 1083  
 QY 997 QDRRDSAMLVARTKPKWENLTSPFLGTLSPS-----ETSEPSILETCALIMKIC 1048  
 DB 1084 QCRSRKGLRAVHNGMOVLSL-----TESPYDGDGHSIQOVSFPLERKAYINVI-- 1137  
 QY 1049 LEIYTVVKSGLDGLKDTLKKEISIEKRFAYWSGYKSLAVHVAETBSGSCSTSLLEYOMLV 1108  
 DB 1138 -----SLKDLITKQORR-----AEVYDSS-----QSHESF 1165  
 QY 1109 SAWRMILLIATTAADIMHLTDSVVRQRLFVLDTGKALLVPAVNCILSGMKTLL 1168  
 DB 1166 SDRKGEILAL-----QOVFLR-----ERSVILAAFRLETLATGTTDAVGL 1207  
 QY 1169 IILRWKRELGSVDEILGPLTEILGV-LOADQOLMEKTKAKVSAFIYVL--QMKEMKV 1225  
 DB 1208 NCLERQIOE-----QVVEYQAAMECLOKDRRLSLBEIOALHQAQNGRKI 1252  
 QY 1226 -----SDLPQSQVLNVCETLQGEVIALFPQTHSLASGATEDKXSMETDDCSRR 1278  
 DB 1253 TLKRESEKSPQELLEVINI--QOKOSQMLEM--QVEISMKRATELQEOISEK----- 1304  
 QY 1279 HRDQDGVCVLGLHAKELCEVEDSDSWLQVTRRLPILPTLLTLEVSJRMK--ONLHFT 1337  
 DB 1305 -----MVVAELSELQA-----TKLELETTLLAQKHKLK 1333

QY	1338	EA-----	TLHLL	TLARTQOGA	NAVAGITG	ISCP	PLSVGLSTNG	TAQT	1384
			:	:	:	:	:	:	
Db	1334	ELEAFRELVK	OTDE	VHLLND	TLASEQ	KKSR	ELQAL	EKEKA	1380
			:	:	:	:	:	:	
QY	1385	PSASRKS	LDASP	WPGV	YRLS	WLS	MEQL	IKTLRY	NFPLDPV
			:	:	:	:	:	:	
Db	1381	EERDK	EELE-----	DLKFS	LESQ	KKQ	RNLQ	LVNL-----	LEQQQLNESQ
			:	:	:	:	:	:	
QY	1444	VQSLAC	LEADHT	VGFI	QLSN	FMKE	WHFHL	POLMRD	QVNL
			:	:	:	:	:	:	
Db	1423	TESQ	RLK	LVDA-----	QLSE	QGR	NLEQL	VLES	EKV
			:	:	:	:	:	:	
QY	1504	HYLQ	NKNGD	GLPS	AVAC	RVCOR	PPS	AA	SPSSKOPAD
			:	:	:	:	:	:	
Db	1470	AQLQ	SSDGTG-----	QSRP	-----	LP	SD	DLK	KBQLQ
			:	:	:	:	:	:	
QY	1564	KT	1565						
		:	:						
Db	1511	ET	1512						

```

RESULT 16
152300      giantin - human
N:Alternate names: gcp372
C:Species: Homo sapiens (man)
C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 09-Jul-2004
C:Accession: 152300
R:Sonda, M.; Mismun, Y.; Fujitwara, T.; Nishicka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized
A:Reference number: 152300, MUID:95100974; PMID:7802676
A:Accession: 152300
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3225 <RES>
C:Cross-references: UNIPROT:Q14789; GB:D25542; NID:9662389; PIDN:BA05025.1; PID:9808865
C:Superfamily: giantin

Query Match      1.7%; Score 151; DB 2; Length 3225;
Beet Local Similarity 18.0%; Pred. No. 0.53;
Matches 332; Conservative 273; Mismatches 661; Indels 574; Gaps 78;

Cy 34  SQT-FAELNKKRRLLBGLSTYKPPSSAAEKVANKKVAAPLKEIGLRISKPLGDBEQ 92
Db 657  SQTLELELNLFH-----KAEIYERKNDKEKANEINLNQLEIE- 693
Cy 93  SVQLQLCYLQEDYRTRD---SVKTVLQDERQSQALLIKINDYYEERTCILRCYHLIT 149
Db 694  -----FKKNADNNSAFTALSEERDQ--LTSQVKE-----LSKVT 726
Cy 150  YFODERHPRYVEYADCVDLKEKELYSKRYQOFEELYKTEAFWTETHGNLTERQVSRWFV 209
Db 727  ELRRNQ-----VKQLENNLAEARQR-----RDYESSQAHMDLILTE-QIHSLSI 769
Cy 210  QCLREQSKLLEITFLYYAVFEMAPSDLLVLTMKERQGFSGSRQTRNRLVDETMDF--VD 267
Db 770  EA-KSKDVKIEVL-----QNELDVQLQFSEQSTLRLRSLOSQLOKNESEVLEGA 818
Cy 268  RIGYFSALLIVEGMNIESLHKCALDDRELEHQFADGLICQ---DMDCMLTFPGDIPHNA 324
Db 819  RVRHRISS-----KVEELSQAL--SQKELEITKVDQLLEKRDVETLQOTT----- 862
Cy 325  PVLIAALLRHTLNPEETSSVYRKIGTAFALQNVQVYTRLLQSLASGNDCTSTMAC 384
Db 863  -----EEKDQOV-----TEISSMTEKKVQVLAEEKFSLG----- 891
Cy 385  VYGLLSFVLTSLDELHTLGNQODIIDTACEVLADPSLPFLFNGTEPTSGIILLDSVCMF 444
Db 892  -----VEITLKEQNLKLSRAEAKKE-----QVEDNEVSSGLKNQNYE----- 931
Cy 445  PHLLSPILLQLRALVSGKSTAKVYSFLDKQSFVNELYKHKPHVDVISHEDGTLNR-ROTP 503

```

Db	932	----	MS-	-----	AAQISKEBEOHFDLLKKENBENRKKLQAALINRELLORVSRLE	979
Qy	504	KULYPLGQGNLRIPOGTV-	GOYMLDDRAVLYRMWEYSYSWTLPTCEIMLI-HVYVTA	560		
Db	980	BEALINXDESKKEIPLSETERGEVEBDE---	NKEXSEKCVTSKCOEILEIYKOTISEK	1035		
Qy	561	DV-IORHOR-----	YKRI-----	IDLXKVIISTDLSDICLLPIT	594	
Db	1036	EVELOHRIKODEEKLAAEBOFQALVXOMOTLODKNOIDLQAEISENOAIIOKL--IT	1093			
Qy	595	SRI-----	YMLORLPTVISPVDVIASVCNCLTVLAANPAAKVMITDLHGTFLPFVAH	648		
Db	1094	SNTASODSDSVAALYKEIVISIP-----	CTGSSEHMK-----	PELEE	1130	
Qy	649	PVSSLSIQ-----	MISAEGM-----	NAGGYNLINMSBOPOG	679	
Db	1131	KILAEKEBEOQOKLOEALTSRAKILKKAQEKERHREELMOQKODRYNROBOPEOSK	1190			
Qy	680	E---YGTIATFLR-ITTLVKGOLGSGSOQ---	GLVPCWFWLKEMLPSTHKRRYNSHG	732		
Db	1191	ENENIGOLROLOQVRESIDGKLPSTDOOSCSSTGLBEPFLKATEOHH---	TOPV	1245		
Qy	733	REQICILBELIHALINLCHETDHSHTSLOPCLICSLATYEAQOTVINIMIGIDVTI	792			
Db	1246	LES-----	NLCPDWPHSHSDASHLO-----	GGTSVAOI-----	1273	
Qy	793	DMVNAAPRSDGAEQOGOLLITVYKLAFSVTNNVIRLPPSNVVSPLBQALSOGHAG	852			
Db	1274	-----	KALQKEIENAE-----	KVELBELKVSSTTSELTKSEBEVYQLOEQINKGO--	1316	
Qy	853	NNLIAVLAKYIYKHGDALPRALIOILIKRLATVAPMSVYACLGNDAAIIDAFLTRIOK	912			
Db	1317	-----	LEBESKTVSHEA-----	EVAHESIQOK	1339	
Qy	913	IEDMRKVMLEPLTVAVETOPGCIELFLNLEVXDSDGSKERSLGWSCSLAHVLEIDS	972			
Db	1340	LESSOLOAGIEHUR--ELOPKDEL-----	OKLISK	1369		
Qy	973	QOODRYNCPPLHRAAIAFLHALMOWDRDSAMVLTRKPKFWENTLSPFLGTLSPSSETS	1032			
Db	1370	KEED-----	VSYSGLQSE--KEAALTGICQIEIIEQEDLICALHTOLOEMQAKEN	1416		
Qy	1033	EPST---LETCALIMKITCLEIYTVVYKSLDLSL-----	KDTLKE--FSIEKRPAYMS	1080		
Db	1417	DERIKOLOVEICEKMOKPEBEIGESSRAKQOIORKLQALISRKALKENKSLQEBELSLAR	1476			
Qy	1081	GYVKSILAVHVAETGSSCTSLIEQMLVSAWPMILLIATIHADIMHUTDSVAVRO-----	1135			
Db	1477	GTERLTYSKLADVSQVSAQNKEDOTVIG--RLALOEBERDKLITTEMDRSLLENOSISS	1534			
Qy	1136	-----LFLDVIDGFKALLVPAVSVCURLGSMKCTILL-----	ILLRQ	1173		
Db	1535	CESLKALBEGITEXEKLY-----	KEISLXSSKIAESTEMOEKHELOKEVEIILLOS	1587		
Qy	1174	WKRELSGVDEILGPLTEILBEGVLQADQOOLMEKTKAVFSAFITVLQNKEMKVSIDIPOYSQ	1233			
Db	1588	YENVSNEAERI-----QHVVEARVROEKQOELVGLKLRSTANKKETEKQLOEAE--OEMEEMKE	1642			
Qy	1234	LYLVANCETLOEVAIALPDORHSLALGSATBD--KDSMEMDDCSRSNHROROGVCYLGL	1291			
Db	1643	KRRFPASKOQKILTELEENDRLAEVHPAGADNAKCEMETLLSSNMAKKELEEVKMEYE	1702			
Qy	1292	HLAKELCEVDDGDSMWQVTRFRLPILPILLTTELEVSILRMQONLHFEATLHLITLARQO	1351			
Db	1703	TLSKKFGOSIMSEKXSLEBYO-----	DLKHQIDGNVSKQANLEATEK--HDNQTNV--TE	1753		
Qy	1352	QGATAVAGAGITGSICTPLLSVYOLSTNGTAQTPSARKSLDAPSWDGVTRLSMLBOL	1411			
Db	1754	EGTOSIPETEBEQ-----	DSLMSSTRPICESVPSAKAN-----	PAVSK--DFFSHDEI	1801	
Qy	1412	LKTRIRYFLPALD-----	PVGQOER-----	TQOCNANARTQOS-----	LACU	1450
Db	1802	-----NNVL-QOIDQLKERIAAGBEEKQXKKEKESQTE--	NEKNVTLISQISTYDGBELKML	1853		

```

Qy 1451 EEAHTYGFILQ-----LSNFM-----EMHHLQMRDIQVNIQVLOAQSTL 1495
Db 1854 QEEVTKNNMLQOQIQEBSRVTKKETAEBEKKDLBERLMMQLAEINQSIGNYQDVTDA 1913
Qy 1496 LHSKMLQHYLQN--KNQDGLPSVAQRVQRPPSAA-----AAPSSSKQPAADTEA 1545
Db 1914 QIKNELLESEKNNLKCVSELEBEKKQDVKEKTVSEIRKEYLEKIQAGQKBPKNKSHA 1973
Qy 1546 SEQQAALTVQYGLLKILSKTLAARHTTPVYCCILLDQSDLDLAENYLPALSTPTPFDS 1605
Db 1974 KELELEKEKOEKOLOK-----DCIRY-----OE 1999
Qy 1606 EVAPSFGLTATVAVNALMGLDLDKXKBPLOAVGLSTQABG--TRUKSLMFMENCFY 1664
Db 2000 KISALERTVKALBFVQGESQKDEITRENNLAQAVEHKKQAQELASPKVLLDDTQ----- 2054
Qy 1665 LLIQANRYLDPRAVHPDRQKQMQEISSELSLSSLSR 1704
Db 2055 ---SEAARVLADNLKKELEQSNSESVKQSQKQKDEDLER 2091

```

```
RESULT 17  
A:Accession: A56539  
      giantin - human  
N:Alternate names: macrogolin  
C:Species: Homo sapiens (man)  
C>Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 09-Jul-2004  
C:Accession: A56539, S37536  
R:Seldis, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.  
Mol. Cell. Biol. 14, 2564-2576, 1994  
A>Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein (  
A:Reference number: A56539; PMID:94187728; PMID:7511208  
A:Accession: A56539  
A:Molecule type: mRNA  
A:Residues: 1-3259 <SEE>  
A:Cross-references: UNIPROT:Q14789; EMBL:X75304; NID:g405714; PIDN:CAAS3052.1; PID:g405714  
C:GeneIDs:  
A:Gene: GDB:GOLGB1; GCP; GCP371  
A:CROSS-references: GDB:454958  
A:Map position: 3q13.31-3q13.31  
C:Superfamily: giantin  
C:Keywords: coiled coil; Golgi apparatus; transmembrane protein  
C:KeyWords: Domain: transmembrane #status predicted <TNN>  
.328-3254/Domin:
```

Query Match 1.7%; Score 151; DB 1; Length 3259;  
Best Local Similarity 18.0%; Pred. No. 0.54;  
Matches 332; Conservative 273; Mismatches 661; Gaps 78;

OY		34	SQI-EAEINCKMRRLBESLSTYYKPPSPSSAKVAVANCDAAPLKEJGRIKFLPGLOBEQ	92
Dd		691	SOIELEINFH-----KAOEITYEKOLBEKAKEISNNLOLEE-	727
OY		93	SVOLLQCYLOEDYRGTRD--SVKTLODEROSQALLIKIADYYEEERTCILRCVLHLTT	149
Dd		728	-----FKKNADNNSSAFALSEERDO--LSTOVE-----LSMWT	760
OY		150	YFODERHPRYVEAVCDVDKLEKEYSKRQOFEEELYKTEAFTWETHGLMTEROVSRWFV	209
Dd		761	EIRKQ-----VKOLEMLTAEAROR-----RIDYSSQTARHNLLITE-QIHSLSI	803
OY		210	QCUREQSMLLETIFLYAYAFEMAPSDLIIVLTMKFEQGFGSGRTNRHLVDETMDP--VD	267
Dd		804	EA-ESKVOKIEVL-----ONEILDVOLOFSESOSTLTRLSOLOCONKESSEVLLEGAE	852
OY		268	RIGYSFALIVEGMIDIESLKCALDDRRRELFHQFADGICQ---DMDCIMTLTFGDIPHA	324
Dd		853	KVRHIIS-----KVELSQAL--SQCELEITTKMDQLLEKRDDVETLOOTI-----	896
OY		325	PVLAMALLRHRTLNPETSVAVRKIIGGTAIQOLNFUYITRLLQSLASGANDCTTSTACMC	384
Dd		897	-----EEKDOGV--TELFSMTKKVNVOLNEKFSG-----	925

QY	385	YUGLSTVLSNLEHTLGNODI1D7ACVULADSLBELFNGTPTGSLGI1D5VCGMF	444
Db	926	-----VEIKTLKEOINLMSRBEAKX-----QVEEDNVSSGLKNYBE----	965
QY	445	PHLSPLLOLRLAVSGSKTAKYFSLDKMSFYNELYKHPHVIDSHEDTLWR-RQTP	503
Db	966	---MSF-----AQGISKEELQHFHFDLKKENEGRKRLQOALNKRKELQVR5BLE	10133
QY	504	KLYPLGGQYTLKRIPOGV--GQYMLDDRVLVWREYSWYTLFTGEIEMLL-HVYSTA	560
Db	1014	BEALNLMDESKKEIPLSETERGEVEEDKE--NKREYSEKCVTSKCOIEIYVLQOTISEK	10693
QY	561	DV-IOHCOR-----VKPR-----IDLNVKYSTDISADCLLPIT	594
Db	1070	EVELOHNRKDLSEKTLABEORQALVKQMNQTLQKTMQIDLLQALQISENQAIIQKL--IT	11227
QY	595	SRI-----YMLQRLTTVISPPVDVIASCVCNCLTVLAARNPAKVTDLRHTGFLPVYAH	648
Db	1128	SNTDASGDSDSVAALKEFTVVISPP-----CTGSSSEHMK-----PELEE	11644
QY	649	PVSSLQ-----MISAEQM-----NAGYGNLNMSSBQPG	679
Db	1165	KILALEKEBOLOKLOLALTSRAIILKKAOEKRNHREELKQOKDDYNRLOEOPDEOSK	12244
QY	680	E---YGVTTIAFLR-ITTLVKGOLSGTOSQ---GLVCFVFLKEMLPZYHKMYNSHGV	732
Db	1225	ENENIGQRLQIQVWESIDGKLPSTDOEGSGSTGLBEPRLPKATENOH-----TQPV	12799
QY	733	REQJGCLLELHAIHLCHETDLHSHST8LQPLCISLAVTAGQTVINIMIGVDTI	792
Db	1280	LES-----NLCPWPSHSDA8ALQ-----G8TSVAOI-----	13077
QY	793	DMVAAOPRSDGEGOGGOLLITVYKLAESVTNNVRLKPPSNVSVPRLEQALSQHANG	852
Db	1308	---KQDLKEIEN-----KVBLELKVSSTSELTKKSEFVQLOEQINKG---	13560
QY	853	NNLIAVLAKTYIYKHPALPRALIQILKRLATVAPMSVYACLGNDAALINDAFILRQSK	912
Db	1351	-----LELESIKTVSHEA-----EVANESLOOK	13733
QY	913	IEDRIKIMLEPFTVAVETOPGLIEFLNLEVKGDSGSKESFLGMSCLHVLLEIDS	972
Db	1374	LESSQLOIAGIEHR--ELQPKLDEL-----QKLSK	1403
QY	973	QOQDRVYCPPLHRAIATFHALWQDRDSAMVLTKPKFWENUTSPLEGTSPSESTS	10322
Db	1404	KEED-----VSYLSGOLSE--KEAALTIKIOTEIIEOBEDILKALHTOLEMOKEH	1450
QY	1033	EPST-----LETQALIMKIICIEIYVYVVGSLDQSL-----QYTLKK-PSIEKRPAYMS	1080
Db	1451	DERIKOULECEMKQKPEEIGESRAKQOIQRLQOALISRKBALEKNSLOEELSLAR	1510
QY	1081	GYVKSILAVHVAETEGSSCTSLLEYOMLVSAMRMLIATTHADIMHLLTDSVVRQ----	11355
Db	1511	GTIRLTKSLADVSGVSAQNKEDVTYG--RLALLQERDKITTEMDSRLLENQSLSS	15668
QY	1136	-----FLDVLVDGKALLVLPASVNCRLSGSMKTLTLL-----TILRO	1173
Db	1569	CESIKIALLEGITBEDKEVLV-----KEISLSSSKIAESTEMOEKKELOKEVEYILLOS	1621
QY	1174	WKRELSGVDEILGPLETILEGVLQADQOOLMEKTKAFYAFITVLQMKEMKVSIDIPQYSQ	1233
Db	1622	YENANSNAEKL-----QHVENAVNQEKELVGKLRSTANKKETEKQOQABE-QEMEMKE	1676
QY	1234	LVLVNVCETLOEVIATFDQTRHSIAGSATED--KDSMETDSCRSRHRDQDGVCVLGL	1291
Db	1677	KMRFTAKSKOQKILIEEENDRRLAEVHPAGDTAKEMETLSSNMAKKELEKRYKMEYE	1736
QY	1292	HLAKELCEVDEBGGSMVQVTRRLPIPLTLLTLEVLSRMQONHFTBATLHLLTLTLARTQ	1351
Db	1737	TLSKKFOPLSNKSDUSLEVO-----DLKHQJEDNVSKQANLEATK--HDNQTNV--TR	1787
QY	1352	QGAATAVAGAITOSICLPLLSVYQLSTNGTAQTPSASRKSUDAPSWGVYRLSNLSMEOL	1411

Db 1788 EGTSGISGETEEQ-----DSLMSRTPCTSESVPASAKSAN-----PAVSK-DRSSHDEI 1835  
Qy 1412 LKTRVFLPEALD-----FVGVHOER-----TLQCLNAVTVQGS-----LAQL 1450  
Db 1836 -----NNYL-QQIDOLKERTIAGLEEEKOKNKEFSCITLE--NEKNTLLSQSTKXDELKML 1887  
Qy 1451 EADHTVGFILQ-----LSNFMK-----BMHFTLPQMLRDIVQNGVLCQACQSL 1495  
Db 1888 QEEVTKNNLMQIQIEELSRVTKLKEITAEEKDXLEERLMOQLMGLNGISGNYCQDVYDA 1947  
Qy 1496 LHSKMLQHYLQ--KNKGDGPSVAQRVQRPSPAS-----AAPSSKQPADTEA 1545  
Db 1948 QIKNELLESEMKNLKCVCSELEEEKQDYKTEKTVESIRKEYLEKIQGAKERGNKSHA 2007  
Qy 1546 SEQALHTVQYGLKILKISKTALRHFTPDVCQILLDQSLDAEYNFLPALSFTTPPDS 1605  
Db 2008 KEIQELKEKQEQEVKQLOK-----DCIRY-----QE 2033  
Qy 1606 EVAPSPFTLLATVNAVLMGLDKEKKEPLTOAVGLSTQAG--TRTLKSLMFTMENCY 1664  
Db 2034 KISLERTVAKLEFVQTESQKDLITKENLQAQAVEHKKQAQELASFVLLDDTQ----- 2088  
Qy 1665 LLISQAMRYLDRPAVHPDKQRMQOELSELSTLSSISR 1704  
Db 2089 -----SEARVLAADNLKKELOSNKESVSKQKQKODEDLER 2125

## RESULT 18

protein T21E12.4 (imported) - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 10-May-2001 #sequence: revision 10-May-2001 #text: change 24-May-2001  
C:Accession: D87755  
R:anonymous, The C. elegans Sequencing Consortium.  
S:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
Science 282, 2012-2018, 1998  
A:Reference number: A75000; MIMD:99066613; PMID:9851916  
A:Note: see webistes genome.wucl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C\_elegans/ published extra appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: D87755  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-4466 <STO>  
A:Cross-references: GB:chr\_1; PIDN:AA837647.1; PID:G1703597; GSPDB:GN00019; CESP:T21E12.  
C:Genetic8:  
A:Gene: T21E12.4  
A:Map position: 1  
C:Superfamily: dynein heavy chain, cytosolic

Query Match 1.74; Score 151; DB 2; Length 4464;  
Best Local Similarity .17.0%; Pred. No. 0.91;  
Matches 283; Conservative 233; Mismatches 533; Indels 614; Gaps 71;

Qy 346 VRKI-----GGTAIQ-----LNVFGYLRLLQSLASGNDCT--TSTACMCYGGLSLF 391  
Db 148 IRTKQLEERDSSGTSIQEMTFMNLERALKITISK--RDEEYVTLLEALKC----- 198  
Qy 392 VLTSLELN-TKGNQD-----IIDTACEVLADPSLPELFWGTPEPTSGIILDSVCG 442  
Db 199 ---GKRHAATGPFSDNGKOKLAVQDYNTLMKEPFLSELVSATDVK-----LMHVVG 251  
Qy 443 MFPPLLSPLQLLRPAVSGSKSTAKVYSELDKMSFYNELVGRKPHDVISHEDGTLMRQT 502  
Db 252 IFLH-----LKRKSTKYPQLRALRLVEAIS-----DLEN 281  
Qy 503 PKLIYPLGGQTNLIPOGTGVQWMLDRAYLVMEYSSWTLFTCEIEMLLHVSTADV 562  
Db 282 SQLKLVSSYVNLKMPPLAEFNEIMSQCALPSKWDDEYDKF-----TALIRD 328  
Qy 563 IOHCORVKEPI-----IDLHVKVISTDLSIADCLPITSRIYMLQ-----RLTTVISPP 611  
Db 329 INKKRRDPSLTKSVKVTAVVHRLFT-----RLMQIILQFRKHQGFRTVIER- 374

Qy 612 VDVIASCNCLTVLAARNPAKWTDLRHHTGFLPFAHVPSSLQ-----MI-SAEGNVAG 665  
Db 375 -----VLRPVGNGSRERQMLDSSG----- 396  
Qy 666 GYGNILANSEPOGEYVTAFLR-----LITTLVKG 697  
Db 397 -----EKSPDQVQVIAVEFLKNVDFLDVDSFAMENAFKRYEDQIGVETALTTLTKS 448  
Qy 698 QLSGSGGLVPCVMFLKEMLPYHKRRYNSHGVQIGGLIELIHAILNLCH--TD 755  
Db 449 QLESRRNS-----NEWFSIFS--RYAALFIRBRIGAIYEYQTRLINRYKEDINE 496  
Qy 756 LHSHTPSLQFLCISLALYEAQGVININGIG-----VDTIMVM 796  
Db 497 LQARFTK-----ARGQGVKIQGTGLPFPSPAKIMIRYERQOLRYMRKREVDVL 546  
Qy 797 AAQPRS--DGAEGGQGLIKTVKLAFSVTNNVIRLKPSPNVVSPLQOALSQHGAGNN 854  
Db 547 GKQENNVDRQQLKADD-----NFKVKNLTQPMFDEWESVQSQWTLPLNK 593  
Qy 855 LIAV-----LAKYIYHKDPLPRLAIOQLRLATVAPMSV----- 890  
Db 594 ILTVDRVQVDRMQLOLKINVHSDSVLYK--EVSHLKSMGFRVPLKTVMAHQANQMRPS 652  
Qy 891 -----YACGNDAAIR--DAFLTRLOSKIEMRIKVMILF-----LT 927  
Db 653 ATSLIEARTFPASVNAALASQVDSLASYKQDIONQLGATVGMDSYKVDYOLKLA 712  
Qy 928 VAVETQGLIEFLNLT-----EYKDSGDSKSEFLGMS 961  
Db 713 ETVTVYQERCEBELLNVRIYNADLNVLKSKRYDKETENLLTSIQKVD--QSLSGNS 769  
Qy 962 CLHVLDELDSQOD-----RYWCPLLRHAAIAFLHLMQDRDSAMLVLRKP 1011  
Db 770 NLAQVNTLDRQETIILARRVEDAIRVW-----TLVFSQSEVELEERQVLPFTV- 820  
Qy 1012 KFMNLSPLFGT-----LSPPSTSRPSTLE-----TCALMKIKICLEYVYVGS 1059  
Db 821 ---GNVAVDLCMTAQTLYISPTRETEKELIQLEYHNSVCTAOMRISGKRFPQVMMDEI 877  
Qy 1106 DQSLKDTL-----KKEFSIERFAPWGSYVSLAVHVAETGSCSTLLEYQML--VSAM 1111  
Db 878 EPEYTHNIIANWMPGQACLEKAYVCVNGIMSDLEEYSE-----WLSQSLMVLQAE 929  
Qy 1112 RMILIIATTAADIMHLDTSVVRQLFTLVDIDGTALLIVPASN----- 1155  
Db 930 QLFEMLGTSLSKMMKTLMERKGRVFDQDTRK--VIFPVSVEYGAQOKILPKYDYWH 987  
Qy 1156 ---CLRQSM---KCTLLIILRWKR--ELGSYDELIGPTELLEGVLDQOOLMERTK 1207  
Db 988 KEMLVKFGAAGVDMQKRFENSVSKRWVLETOYD--GGSSTPTIGLISVQSLKKQTK 1044  
Qy 1208 AKVPSAFTVYLQMKEMKVSIPQYQSLVNLNVCETIQEVEVIALFPQTRHSLALGSATEDKD 1267  
Db 1045 S-----GQAVDLYRSSQRLN-----QQRXQFP----- 1068  
Qy 1268 SMETDSCRSRHRDQDGVLCVGLHLAKELCEVEDGDSMLQVTRRLPILPTLLTLEVS 1327  
Db 1069 -----AQMLYSENVBE--MSAFTEILSRDASIQIQMMN 1101  
Qy 1328 LRMQNHLFTREATIHLILLTLARTQOGAT-----AVAGAGITQSGICPLSVYQSLSTNGT 1381  
Db 1102 LQTK-----FAQED--ELVEKRYVETLLENKRSKPVGAQRPQ--ALNVITAFKAKNL 1153  
Qy 1382 AQPSASRK-----SLDAPSPG--VYRLSMSLEBOLTLR--YNFLPEALD-- 1425  
Db 1154 TREBNRKRRKARVALDLDSDAHAPBSGDKLYATATELAAKMDVWALQPVYTGIDBAEKT 1213  
Qy 1426 FVGVHOERTIQLCLNAV-----RTVQSLACLEADHTVGFILQ-----SNFMKE 1469  
Db 1214 WLSVQPKRKIQSLDELANNQQLPVCRCRTYKSVBHVQMHLTYGKMMVLVAELKESALKE 1273

QY 1470 WEFHLPOLMRDIQVN-----LGYLCOACTSLHSRKLQHYLQNKKGDLPSAVARV 1522  
 DB 1274 RHMH--QWKEKRVVMNLSDLTLGQVMD-----DLRHE----- 1306  
 QY 1523 QRPBSAASABSSKOPAADTASEQALHTVOYLKILSKTIALAHRFPDYCOILLD 1582  
 DB 1307 -----HTIKKILL--VAQGEWALBEFLREMEYQON 1335  
 QY 1583 QSLDLAEY-----NPLFALSTPTTFD--SEVAPSGTILATVN 1619  
 DB 1336 YEVELVYONKTRILKGMDDLKKEHONSLSAKKL--SPYKQPESSAQSMDEKLNKIN 1394  
 QY 1620 VALNMLGELDKKKEPLTOAVGLSTQAEGRILKSLMFTMENCEFYLLISQAMRYLDPAY 1679  
 DB 1395 AMFVDWIDVOR-----WVYLEGLFSGSAEISTLLPFESSR--PATITTDVLMKKVAA 1447  
 QY 1680 HPR-----DKRMKELSELSTLSSLSRYFRGAPSSP 1714  
 DB 1448 SPRLIDVNVMOGACRLRLADMLAKIQKALGEYLERESSFP 1490

## RESULT 19

119493  
 hypochetrical protein C26H9A.2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C:Accession: T19493  
 R:Smyle, R.  
 submitted to the EMBL Data Library, September 1997  
 A:Reference number: Z19131  
 A:Accession: T19493  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1334 <WTL>  
 A:Cross-references: UNIPROT:Q9YTS7; EMBL:Z99169; P1DN:CAB16307.1; GSPDB:GN00022; CESP:C2  
 C:Genetic: A:Experimental source: clone C26H9A  
 A:Gene: CESP:C26H9A.2  
 A:Map position: 4  
 A:Introns: 135/2; 208/2; 301/3; 329/2; 396/3; 469/2; 504/3; 588/1; 619/3; 638/3;

Query Match 1.7%; Score 149.5; DB 2; Length 1334;  
 Best Local Similarity 18.1%; Pred. No. 0.16;  
 Matches 299; Conservative 203; Mismatches 530; Indels 621; Gaps 76;

QY 236 LVLVTKMFK-----QFGSRQTNRLVDETMDFVDRIQYFSAITLVGMDIESLHKC 289  
 DB 11 LHLRLKTFSEYLKIPVSGSRSDPSRL-----PLFHKVSNSWIF-----FHKV 55  
 QY 290 ALDDRRLHQPAODGLICODMDC-----LMLTFGDIPIHNAFVLLAMALLRHTLNPEETS 344  
 DB 56 FINN-----IPQVMSMTTPOQLNAEFREVVHFTFL-----CSV 89  
 QY 345 VVERIGGTALQNVFOYLITRLQSLASGNDCTTSTACMCVGLSFLVTSLEHTLGNQ 404  
 DB 90 LKVEVQRAASTGTIEAQSIAEFLRG-----TELKGTIIEAIRFILSS-----ED 137  
 QY 405 QDIIIDACEVLAPSPBELFWGTBPSGLGIIIDSCVGMFPHLLSPDLLRLALVSGKST 464  
 DB 138 EIMIDAACKV---SLPSTLVKT-----YLFDP-----LPAATTVVTDL 175  
 QY 465 AKKVSFLDKMSFYNELYKHKPHDVISHEDEGLIMRQTPKLLVPLGGQNLRIPOQTVO 524  
 DB 176 TENATETBEMLKNNE---KLHDMT-----MLHNVSTADYIOH-----CORV 569  
 QY 525 VMLDRAVLVWREYSVSWTLFTCEIE-----MLHNVSTADYIOH-----CORV 569  
 DB 200 IM-----EGLCRFKCVSEELVRKODLLLFVGTTQVAAENNMCKRLCARL 245  
 QY 570 KPIIDLVHKVISTDLSIADCLLPITSRIMLQRLTIVSPVDVASCVNCCLITLAARN 629  
 DB 246 LEVI-----ADKSTINAVIKTV-----IVKCCVRIFVKNLSGA 278

QY 630 PAKVITDLRTG-----FLPVAHPVSSLSQMIISAEQMANAGY---GNLLMNSQPO 678  
 DB 279 PSHSHSDAQTAHSIICLSCEFLKNSAY-----LTDQLDEFEADGYIVIKFLKSPT-- 332  
 QY 679 GEYGVITAFRLITTLV---KGLGSGTQSGVLPCWPFVUKMLPS----- 721  
 DB 333 -DNEIYVNIILMISLVNSKKEFTTPQYISGLVQLPAFOLP--TPANGSLSVNLAAPAL 389  
 QY 722 -YHKRYNSHGVBEQIGCLIELIHAI----- 747  
 DB 390 LYHVF---LESTSESTSCVYIDILHSIYTCODPANYFILDKEYPELSIFIDQLEKHPHAVY 446  
 QY 748 --LNLCHETDHSHTPSIQFLCISLATE--AGQ---TYVNIIGVDTIDMMAAO 799  
 DB 447 KLIELIEPAVFOUSHIPCEBLSLCVLLKTEISAGTACTTIYQWCFPLITDAIIXKA 506  
 QY 800 PRSDG-----ABGQSGOL--LIKTVKLAFSVTN 827  
 DB 507 FREVGLDLALCFIIRMFEMYEKNIQNASFTVSTSSRQDQKLSLTTDLITTIKKN 566  
 QY 828 VIRKPSNVSVSLEQALSOHG-----AHGNMIAVAKYIYHKHDPALPRLAIQLKRL 882  
 DB 567 -----TENGLFTECPGAKLLEIYCEVNEBWRSSL---LQLVKOL 604  
 QY 883 ATVAPMSVYACLGNDAAIRDAFLTRLOSIEDMRKXWILBELVAVETQGLIELFLN 942  
 DB 605 LIVAF-----TDQYIMLVNTDQD-RPIENLEVMFSLKALLGVVRESHK 649  
 QY 943 LEVQSGDSKESPSLGMSCLHAYLELDSQOQDRYCPPLHRAA-----IAF 991  
 DB 650 VRIQFRKSGG-----FLALSALLGLEKRPFD--LHPAGDGRIPREQOMLEF 695  
 QY 992 LHALMDRDSAMVLRTEK-----FENILTSPL-----FGLSPSTSEBS 1035  
 DB 696 VHIIFR---VFTLSRREPSNAKFTSEITWDSITSLRLTSIPNESIVSIEBTEWKS 751  
 QY 1036 ILETGALIMKICLEIYVYVKGSLDQSLKDTLKKFSIEKR---AYMSGYVSLAVHVAE 1092  
 DB 752 LHTHDLASEIACHVVF-----RIDDNEAGNVKGMFNIYFGCYCRLLFNNA- 801  
 QY 1093 TEGSGSTSLLEYQMLVSMKMLIITATTHADIMHTDSVVRQLELDVLDGKALLVPA 1152  
 DB 802 -----LDNYEKNTSDIRW-----NDGASLESISIV-----SWTSSILVHG 837  
 QY 1153 SVNCLRLGSMKCTLLILLRLQWKEIGSVDELIGFTILEGVALQADQOLEKTAKFPS 1212  
 DB 838 AII-----SMCLPESIASASKWTIAQYVSLKAILKS--BRNOQIMCOV----- 884  
 QY 1213 AFTVLQKMKVSDIPQYSQVLVNVCT--LOEVIAL--FDQTRHSIALGSATED--K 1266  
 DB 885 -----DMPKH---LRIAEKLFLENHILLOFPYULLERLTSQSLTPNQLR 927  
 QY 1267 DSMETDD--GSRSRRRDQDVCVGLHLAKELCEVBDGSMLOVT-----RPLP 1315  
 DB 928 SFLRLDSPLCKSLDDDE-----DDDDSKIEIADNEGPPVPLQHVK 969  
 QY 1316 ILPILLTLEVLSRWKQMLHFTFATLHLLTLARTQCATVAVAGITQSIPLPLSYQ 1375  
 DB 970 ALVSWMT-----PDDQYIGTA-----PSVPEFD 992  
 QY 1376 LSTNGTAOTPSASKSLDAPSPGVYRLSMLFOLKTLR--YNFLP---BALD--FVG 1428  
 DB 993 MSVEGFA-----ALYLPISLAPMF--STTKARIEPPLNGSFSCWYIYDALSDKAD 1042  
 QY 1429 VHOERTQCLNAVATVOS-----LACQ-----BEADTVGFIL 1461  
 DB 1043 AHPRLILTVTAVAAPNSDPSKSPSQTALHACQCPASAFDRSLISTEBSDQPGADLE 1102  
 QY 1462 QLSNFMK-----EWHFHP---QLMRDIQVNLGYLCOACTSLHSRKM 1501  
 DB 1103 KTAHFOTDKLIRTLADQVRCGEW--FLHAVVFNRSVLAKSGVSL--YL-----NGHISTQK 1156  
 QY 1502 LGHYLQNKNGDGLPSAQAQVQRPASAA--APSSKOPAADTASEQALHTVQYGLL 1559



Db 1157 LMTVAQAGAAQGLAQGVSNVAVGTLPLRLRSRLRFLASFLAE----- 1205  
 Qy 1560 KIISKTLALRHFTPDVQCILLDSLDLAEYFLFALSTFTPTDSSVADSGTLATVN 1619  
 Db 1206 PMAEFTVRAVAVQLP-----HYIGNYQASTAPLPEEKIVFSLSAATOE 1251  
 Qy 1620 VAL-----NMGEIDKKEPELTQAVGLSTQAECT 1648  
 Db 1252 LTLAKIRTMYGKMD-AEILSQHLGISPDHST 1282

RESULT 20  
 T14156  
 kinesin-related protein - African clawed frog  
 C/Species: Xenopus laevis (African clawed frog)  
 C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
 C/Accession: T14156  
 R/Mod: K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.  
 A>Title: CENP-E is a plus end-directed kinetochore motor required for metaphase chromosome  
 A/Reference number: 217893; MUID:98028574; PMID:9363944  
 A/Accession: T14156  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-2954 <MOO>  
 A/Cross-references: UNIPROT:O42263; EMBL:AF027728; NID:g2586070; PID:g2586071; PIDN:AACC  
 C/Genetic: XCENP-E  
 A/Superfamily: centromere protein E; kinesin motor domain homology

Query Match 1.6%; Score 147; DB 2; Length 2954;  
 Best Local Similarity 17.1%; Pred. No. 0.88; Mismatches 474; Indels 508; Gaps 57;

Matches 244; Conservative 202; Mismatches 474; Indels 508; Gaps 57;

Qy 557 VSTADVYQHCORVPII-----DLVHKVISTDLADCLPI-----TSRI 597  
 Db 1 MSBGDAVKCVVRPFLQREGDQANLQKAGNTTQVQDTSFNDVFNHSHSTQI 60  
 Qy 598 YMLQRLTVISPPVDVIASCV---NCLTVLAARNPAKWTDLRHGTFLPFAHVPSSL 653  
 Db 61 YQET-----AVPIRSLQGYNGTIFAVGQTSKGTYYMGTNSLGIIPQALQEV 111  
 Qy 654 SQMISAGNNAAGCGNLMSSEPOGEGVTIARLITLVKQL----- 699  
 Db 112 FKII-----QELPFRREFLRVSYMEIYNETVXDLLDDRRKKPLEIRE 155  
 Qy 700 -----GSTOSQGLVP-CVMFVLK-EMLPYHKRVNSHGVRQIGCLIELHAIL 748  
 Db 156 FNRNVYVADLTBELVMVPEHYIQIKGGEKRRHYGETRMDHSRS-----HTIF 205  
 Qy 749 NLCHETDLHSHRPSLOFLCISLAYTEAGQTVINIGIVDTIDWMAAQRSDAGEQ 808  
 Db 206 RMIVESHDRNDPTSEN-----CDGA-----VWVSHNLVDLAGSEBSAQTAEG- 250  
 Qy 809 GQGQGLIKYKLAFSVNNVIRLKPNS-----VSPRLBALSHQHAGNLLAVLAKY 862  
 Db 251 -----VRLKEGGINRSFLILQGVIKLSDGAG-----F 281  
 Qy 863 IYHKHDPALPRLAIQLKR-----LATVAPMSVYACI-----GNDAAAIR- 902  
 Db 282 INNR-DSKLRILONSIGNAKATVITCTIPVSDFTLSTQFPAKAYRNTPHVNEVL 340  
 Qy 903 -DAFLTRLOSKIEDMRKIMILEFLTVAVETQGLIELFLNIEVKGSDGSKESLGMW 960  
 Db 341 DDEKLLKRYREKIDLLKKOLENLE-----SSSETKAQAMAKESHTQ--- 381  
 Qy 961 SCHAVLELDSQODRYWCPRLHRAAIAFLHLMODRPSAMLYLRTKPKWE----- 1015  
 Db 382 --LTAELKQHKEREDRW--HLTNIVASSQSQDQD-----VRRKRVYTWAPSKIQ 431  
 Qy 1016 -----NLSPFLGTLSPSETSE-PSILETCALIMKILCLE----- 1050

Db 432 NSLHAGSVDFDMLSRPLGNFSKAKFSDMPSFPE-----IDDSVCTEFDFD DALSNMDS 487  
 Qy 1051 -----IYYVKGSLDQSLKD-----TLK 1068  
 Db 488 NGIDAEWNILASKYTHREKTSLHOSMIDFGQISDSVQPHSSKENQOLYLPKDSGMACE 547  
 Qy 1069 KFSIERKFAVY-----SGYKSLAVHVAETGSSCTSLLEQMLVSAVRMLLI 1117  
 Db 548 KASFEKITSLOOOLSKSEKEKELVOSFELKIALEBQSLSVAKULEMTNS-REHSIN 606  
 Qy 1118 ATTHADIMHLDTSVRRQLFDVLDGFKALL-LVPASVNCIRGSM-KCTILLILLRQ 1173  
 Db 607 AVQTDV---EKVVRKEMSVLDGSGYNASNDLQDSVQKRLSSHDCEIHRKMLEQ 663  
 Qy 1174 WKRELSGVDEILGPIREILGVIQAO-----QIMETKAKVFAPIITVLQMKMKV 1225  
 Db 664 ---KIVLEEFINLNKSSKENDQKSSQSDPMESIQCEAIMKALBELALMRD--- 717  
 Qy 1226 SDIPQYSQVLANCETLOEEVIALFDQTRSLALGSATEDKDSMETDCCSRSHRDQDQ 1285  
 Db 718 ---NFDNIILE-NETLKREIADL---ERSL-----KENQETHEPELEKETQEH 760  
 Qy 1286 VCVLGLHLAKELCEVEDGDSWLD-VTRRLPIPLITLLTLEVSU---RMKQMLHTEAT 1340  
 Db 761 EAQL-IHEIGSLKKLVENAEVYNQNLDEDELETYKTLKEQEIQLAELRKADNLQKKVRN 819  
 Qy 1341 LHLTLTARQOQATVAVAGGITQSTCLPLSYQSLSTNGTAQTPGASR----- 1389  
 Db 820 FDLVSN-----GDSKLCSEEIFOLKQSLSDAEAVTRDAQKCSFLRSNLE 866  
 Qy 1390 ---KSLDAPSPGVRRLSMELDKTLRYNF----- 1419  
 Db 867 LKEMEDTSNWNQKEKASLPFKOLETSKNYKMAADLOKELOSAFNEINYLINGLAG 926  
 Qy 1420 -----LPEALDFVGHQERTQICINAVRV-OSLACL----- 1450  
 Db 927 KVPRLDLSRVELEKVSFESQKLEKALEKNA-LENEVCTISEKYPLENEVECKNOISK 985  
 Qy 1451 -----EADHTVGFILQLSNPKEMHFLPQLMRDQVNLGVCOACTSLHSRKM 1501  
 Db 986 ASEBIMLKQEGHSASISIKOELIMOQSOQLQLDVETHROSKVQOQEOYLEMKM 1045  
 Qy 1502 -----LOHYLQNK-----NGDGLPSAVQORVQRPASASAAVSSKQPAADTEAS 1546  
 Db 1046 HDLFEKYLNNKBAEDLLAEMENLKGTMESVEKI-----ADTKHELEERTIRD 1094  
 Qy 1547 EQQALHTVOYGLKIISKTLAALRHFTPDVQCILLDSL-----DLAEYNF 1592  
 Db 1095 KEQDLHEKTY-----FQAMQTIPIPIR-----LSDSLPSKLVGNSQDPIEINDYHN 1143  
 Qy 1593 LPLSFTTPFPDSVAPSGTLATVVALNMLGELDKKPELTQAVGLSTQAEGRITLK 1652  
 Db 1144 LIAL-----ATERNNIMVCELETRNSIKE-----QVIDLMTQ----- 1175  
 Qy 1653 SLMTFTWENCFYLLISQAMRYLRDPAVPRDKORMKQEL-SSSELSTLL 1699  
 Db 1176 -----LQSLQKQSIKSDLDQKPDLEBGEVKLL 1205

RESULT 21  
 S48938  
 hypothetical protein YHL030w - yeast (Saccharomyces cerevisiae)  
 C/Species: Saccharomyces cerevisiae  
 C/Date: 02-Dec-1994 #sequence\_revision 02-Dec-1994 #text\_change 09-Jul-2004  
 C/Accession: S48938  
 R/Favella, T.  
 submitted to the EMBL Data Library, June 1994  
 A/Description: The sequence of S. cerevisiae cosmid 9196.  
 A/Reference number: S48794  
 A/Accession: S48938  
 A/Molecule type: DNA  
 A/Residues: 1-1868 <FAV>  
 A/Cross-references: UNIPROT:P36737; EMBL:U11583; NID:g2269854; PIDN:AB65042.1; PID:g2269854



C:Genetics:  
A:Gene: SGD:ECM29  
A:Cross-reference: SGD:S0001022; MIPS:YHL030w  
A:Map position: 8L  
C:Keywords: transmembrane protein

Query Match 1.6%; Score 146.5; DB 2; Length 1868;  
Best Local Similarity 18.2%; Pred. No. 0.45; Mismatches 632; Indels 609; Gaps 92;  
Matches 337; Conservative 269; Mismatches 632; Indels 609; Gaps 92;

QY 61 SAEYKANKDVASPLKELGRLSKFLGLDEBQSVQLQCYLQEDYKRGTRSVKTVLQDER 120  
DB STDSTNLSDDAAQFMKRP-----MPENEEFITFLQTLVGTANGRPVKAILQEKI 322  
QY 121 QSQALLKIDAYYEEETCI-LRCVHLITTPQDEBHPYREYADCYDKLEKELVSKKQ 179  
DB 323 LS--ILNRSHFAVTKKACISLISIGL-----HSSEY-----KASLTLSFIRH 364  
QY 180 QFEELYTEAPMTWETHGNLMTERRQSRWFVQCLREQMLLEIFLY--VAFEMAPSDL 236  
DB 365 VAKINYKINLPA-----SSPSSTPSTCIYSLINN-----LHAGMFKQLGP-- 409  
QY 237 LVITKMFKEQFGSGRQTNRLVDETMDFVDR-IGYFSAIIVBGMDESIHAKCALDR 295  
DB 410 -----QTPAFNTAIIQRLQYETLGDILKRDDELVSLSYIEFL-FESL-----KN 454  
QY 296 ELHQFADGILCOMDCLMTFGDIPHHAPVLAMALLRHTLNBERSSVVRKIGTALQ 355  
DB 455 DLPPFRSS--IQESLSLVGHLSILPOOSKIKLK-NILRKLSIDEQO----- 499  
QY 356 LNVQVYTRLLQSLASGNDCTTSTACICVGLSPVLTSLDELHTLGNODIIDTACEVL 415  
DB 500 -----REDNNDAVNSIMAL-KFVSIKFTNNAAPRH----- 528  
QY 416 ADPSLPFLF-WGTEPTSGILIIDSVCGMFP-----HL- 447  
DB 529 -DPE-ARLFNMGVTRNRPDIIEESFKGLQPFWRVNNASINTSATVKTSDLLGSHISE 586  
QY 448 --LSPLQLRLALVS-----GKSTAKKVSFLDKMSFYVELYGHKHVDVSHEDGT 496  
DB 587 TEPPFPFEPQLVLDQDSEAASTTRKSLNNAVAFSQCLISNAYGKTM-VIQDED-- 643  
QY 497 LMRQRTKLLYPLGGQTNLRIPQGTQGVMLD--RAYLVREWSYASWT----- 544  
DB 644 -MSVRIDKAL-ELDDTVASRVNEWVG--NMDDIFIRYLLTSLNEPATNSKGQIAIFP 639  
QY 545 ----LFCSELMLHVVSTADVIQHCORVPIIDLVAKVI-----STLSLADCLLP 593  
DB 700 YDPIFGSVLLTLNLFVSN-NVLRRLLEIVP--DLVHLVIMKQSLSDNDLAVCATIIGI 756  
QY 594 TS-----RIVMLLQRLT--TVISPPV-----DVIASCVNCLTVL 625  
DB 757 ISTRALDSTHVKRTKIAQSGTMAETVYASVVRRLYLKQTNHIESDSLNLNLITL-- 814  
QY 626 AARNPAKVTDLRHTGFLPVAAHVSSLSQMSISAGNAGGYGNLLNNSQOPQEGYVTI 685  
DB 815 -----THLSHPG-----TNKDMILKLVQCVTKFGILLQVSAQERKD----- 850  
QY 686 AFLRLITLVKQGQSGOSQGLVPCVMFVLEKMLPSYHKRRYNSHGREGICLIELIH 745  
DB 851 -FLKRVMDTIQDKLIND-----VTALQTSVLS----- 877  
QY 746 AINTLCHETDLHSSHTPSLQFLCISLAYTEAGTVININGIVDTIDWMAAQPSRSDA 805  
DB 878 -----LYSTDLENS-----SLFQKLELTNVSQKN-----DLFSYGBESLSV 915  
QY 806 EGGGQGGQLIKITYKLA-FSYTNNVIRLK-PPSNVVSPLQALSQHGAGNNLAVLAKYI 863  
DB 916 AGKWSKYLKQIDIPFNV--EIMQOKFPATVNTITLD-----1 954  
QY 864 YHGHDPALPRL--AIQLKRLATVAPMGVYACLGDA-----AATPDAFLTRLQ-- 910  
DB 955 FSGCDSTKPSLRKASCIWLISYQYIGHLPVSSKCNIDHLRMRPLADDEFIQDSAPR 1014

QY 911 -----SKEDNRKIMILEFL-----TVAVT-----QPGILFL-- 939  
DB 1015 GLSLVIEIGSDDKESVWKGILKSFTESTAGSASTATGISGVSEETLEFEPGLVMTGD 1074  
QY 940 -----PLNL--EYKDGSDGSKESPL-----QWMSQ-----LHAYL-----ELIDS 972  
DB 1075 GSISTYDILNLASEVDPLVYTKFMSLASALMSRSKRIAGLGAIMSKSSILEELLK 1134  
QY 973 QQGDYWCPELLRAALAFPLHALMODRDSAMLVLRTKPRFW-----ENLTSPLF----- 1022  
DB 1135 DQO-----TAKKILPKLYRFRFPQAVSRMTDIIWMTLIPESLITSLYENDI 1183  
QY 1023 -----GTSPSEBSEPSLFTCALIMK-----ICLEIYVVGSLDQSLKDT 1066  
DB 1184 LDELICGMANKRVRBAS--TSALLQLIQSOPQOEKPSKMLKIWTMARFTMD-DIKDS 1239  
QY 1067 LKFSIEKRPAYSGVYKSLAVHVAETEGSSCTSLLEY--QMLVSAM-----RML 1114  
DB 1240 VREVGTR-----FTYVAKILARSIDVEKGVNPTKSEILDNIILPFLMGPHGLNSDAEVR 1295  
QY 1115 LIIATTHADIMLTDVVR--RQLFLVDLGTAKALLVPASVNCRLGSMKCTLLIIL 1171  
DB 1296 NPAITLIDLVKSPGAIKFPFKLYDFI--TLFSIEPQVINYALNANANTIDANVI 1353  
QY 1172 RQKRELGSDVDELIGLTELLEGVL-QADQQLMEK-----TKAYFSAFIT 1216  
DB 1354 -DTQRKNGVTN--SPFQTEIKLNNSDCMMEIINVIKASRKSQGLSPKVASSLVI 1409  
QY 1217 VLQMKEMKVSIDIPOYSQVLNVNCEITLOE-----VIALPQTRHSLA 1258  
DB 1410 IILVKKYSI-EMKPYGKILKVLNMFEDRNEVSNIAPALSMGYLERVSLDCKIKYSEK 1468  
QY 1259 LGSATDKDSMETD-----DCSRSRHDDQGVYGLGHLAKELCEVEDGD----- 1305  
DB 1469 LITKYPEPTSTENKKRVGTALDISILNVAKSEEDNVA5VFMPIFLACN-DEKQDLETLY 1527  
QY 1306 --SMLOV-----TRRLPILPFLTLLEVSILRKQKHLHFEATLHLLTLARPOGATNV 1357  
DB 1528 NKTWTAASSGACTVCL-YLPEILANVLCVNIKSN-----FSIRKCAKASVIO 1574  
QY 1358 AGAGITQSICLP-LLSVYQSLSTNGTAQTPSARSKSLDABSPG--VYRLSMSLMQLK 1413  
DB 1575 LCGINDSIPIYPQIVKLPDI-----SEALSGRSMQCKEIHIVALVSLTEKFSQ 1623  
QY 1414 TLR-VNPLPEALDPVGHQERTLOCUNAVRTVQSLACLEADHTVGFIIQLSNPKEMH 1472  
DB 1624 TVADNNDLQESINHV-MYTEVSRSKMYVYKXI----- 1654  
QY 1473 HLEPMLRDIOVNLGVLCQACTSLHSRKMLOHLYOKNGDGLPSAVALRVQRPASAAS 1532  
DB 1655 -LPLVARYINNVQF--EITFLIEKAKEMIRLIGSESDS-BGSLKQ-----TSD 1701  
QY 1533 PSSSKQPADTEASEQOALHTVQ-----YGLKILISKTLAALNHTPTDVC-----QI 1579  
DB 1702 ESTIKRIKENTETQSGSKENIENEYVNLKV-----SVDICNNSKSRYPM 1750  
QY 1580 LLDQSLDLAENVFLPALSTPTTPDSEVAPSGTILATVVALNMG 1626  
DB 1751 LLEFIID-EIAYLF-----HNDRIIHTWRTQLAASEIGISIVG 1787

RESULT 22  
T32550  
hypothetical protein f33b4.2a - Caenorhabditis elegans  
C/Spectes: Caenorhabditis elegans  
C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
C/Accession: T32550  
R/Johnson, D.; Stellyes, L.  
submitted to the EMBL Data Library, December 1997  
A/Description: The sequence of C. elegans cosmid f33b4.  
A/Reference number: Z21190  
A/Accession: T32550

A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2848 <JOH>  
A:Cross-references: EMBL:AF036702; PDB:1AB8374.1; GSPDB:GN00022; CESP:F33D4.2a  
A:Experimental source: strain Bristol N2; clone F33D4  
C:Genetics:  
A:Gene: CESP:F33D4.2a  
A:Map position: 4  
A:Insertions: 48/2; 117/3; 167/3; 263/3; 316/3; 362/1; 378/2; 396/1; 436/3; 509/3; 658/3; 7  
/1; 2324/3; 2430/1; 2492/3; 2597/1; 2823/3  
C:Superfamily: Inositol-trisphosphate receptor

Query Match 1.6%, Score 145.5; DB 2; Length 2848;  
Best local similarity 17.0%; Pred. No. 1;  
Matches 325; Conservative 264; Mismatches 653; Indels 669; Gaps 84;

QY 9 SVLSFCSSRELWTI--LIGRSALRELSQIEALNKMRRLLEGLSYKKPPSPSAEKVK 66  
DB 851 SPMSAIHARLMMISIPENVNVSTYESVS-VEASDGSRMRIEGE-----AHKVL 899  
QY 67 ANKVASPIKELGRIKFLGLDEOSVQLQCYLQEDYRGRSVKTVLQDEROSQALI 126  
DB 900 ATYETV---WGLRNO---SMEERQSV-----NSSKLTVEIVNLAKL- 936  
QY 127 LKIDAYVEERTCILRCVHLHLYFOD---ERHPRYEYADCYDKLEKELV-SKYRQOF 181  
DB 937 ---AQFHFSPNDLQLTQNLALINEGRATEQVPSHRAMVNAIRNMSKMRGKNENS 993  
QY 182 BELYTEAPFETHG-----NLMTEROYSR--WPVQCLREOSMLEIIFLYAAVFE 231  
DB 994 KDLAKTSPVAEEAGRTKEGRALNVKTLIVAEILQFVMDVRDRIYITMALSPKXNF-- 1051  
QY 232 APSDLVLTKMFKKQ--FGSRQNRHLYDPTMPPVDRIQYFALLIVEGMDLESLHKC 289  
DB 1052 -PC-----DGLSMHSASINERASELYDAIYRSSG----- 1083  
QY 290 ALDRREHOPADQGLICQMDCLMTFGDIPHAAYVLAALLRHLNPESTSVYRKI 349  
DB 1084 ---ELHIDGRGQGLLAI-LLOMTMSDVPPLTSTALK-VFRHRTQYELLEDKQV 1135  
QY 350 GGTAIQLVNFOYLTRLQSLASGNDCTSTACMCVYGLLS--FVLTS--ELHTLG 402  
DB 1136 -----QLVSNNDVEN-----YQIDRDIFILNLEKSELWVHG 1170  
QY 403 NQODIITACEV-----LADPSL--PELFWGTEPTSGIILIDSVCGMPPHLSR 450  
DB 1171 DRHHSIDTK-EVDEKERTTEHLLDHDLSPRADSGSMELAAVLNE--HYPSIRNE 1226  
QY 451 LLOLLRALV---SGKSTAKKVSFLDKMSFYNELYKPKPHDVISHEGDTLWRQTPKLLY 507  
DB 1227 CLQALNLLIKDNDNAVALQELSDK-----APLIAY 1259  
QY 508 PLGGQTNLRI-----PGCTGVQVMDRAVYVREWEYSYSWTLFTCELEMLHV 556  
DB 1260 PLIKQMLVRLTGMCKYRKGDPRKPTMNOQLTKMRVY-----EVLLEF 1301  
QY 557 VSTRADVQHCORCVPIIDLVKAVSTDLSTADCLPITSRTIYMLQRLTYISPPVAVIA 616  
DB 1302 ISVPHDKHDHMKLITLISHEFLR--SFCKTKENOSRLY----- 1340  
QY 617 SCVNCITVLAAARNPAKWTDLRHGTGLPFVNAHPVSSISOMISAEQNAAGVGNLNNSEQ 676  
DB 1341 -----KFIYEKDAEGMLRVETIE----- 1360  
QY 677 PQGEYVITAFRLITTLVKQGLSGTOSQGLVPCVMFVLKEMLRSHKMRVNSHVEQI 736  
DB 1361 ---EVGTLVAIFR-----NNRELASNVPEELIAHVGLE-----HNSRNP 1399  
QY 737 GCULLEIHAALNCHETDLSHTPSLQFLCISLAVTEAGQVIVINIMIGVUTIDMW 796  
DB 1400 ---FLELLQALV-CVYKEIESGQEKAVNEICAAS--DEVRLQYVD--NASFEELAM 1450  
QY 797 AAGPRSDGAGGQGGOL-----LIKTVLAFSVYNNVYRLKPSNVVSPLEQALSQGH 851

DB 1451 KDEKSKGRSSDSRRKLYHIELVRLAMCTRGNGTELKASQI--PMD----- 1499  
QY 852 GNNLIATLAKTYHKKHPALPRLAIQLLXKLATAPSPSVACGLNDAAALRDAPLTR--- 908  
DB 1500 ---HIVRVVT-----AKQCLVKKYTYQLLHHCYDTPAEKMDAYKTEYVD 1543  
QY 909 ---LQSKTEDRIKVMILEFTVAVERQPLIELFLNLEVVDGSDGSKFSIGMSCLHAV 966  
DB 1544 HILNMLLED--IRSLREKLT-----GATATLEHIC-HIV 1577  
QY 967 LELIDSOQDQRYMCPRLHRAAIAFLHALWQDRDSAMLYLRTPKFWMILTSPDFGLS 1026  
DB 1578 TEVL-----IKFPAFYSALQAAVVDVHHKKTFSE----- 1608  
QY 1027 PPSSETSPLTECALIMKIICLETIYVVGSLDQ-----LKDTLKF----- 1070  
DB 1609 -----VLELTYLEKGRKSGSKSSHWYRVACIRKLTGMAEHN 1648  
QY 1071 -----SIEKRF-----AYMGVYKSL-----AVHAFTEG 1095  
DB 1649 ITLPATLAGPOMSGQTSVRQWQOASAKWIGIGKRLNRONTLNPGRHLYGTSNPTHE 1708  
QY 1096 SSTSLLEYOMLVSAWMLLIATTA-----DINH-----LTDVVRQQLP 1137  
DB 1709 TSANVVTGYHMMIGEFKFL--HPLHAAGSVLVEVHLTELLFPEGASALRDQARGGV 1766  
QY 1138 LDVIDGTKALLVPASVNCIRLGSMTCTLLILIRQKRELGSVDILGPLY---FILE 1193  
DB 1767 AKLIQCKTLMONKODNICARVLOTLCMKDCTKQ-----LTHGHRVYN 1812  
QY 1194 GYLQADQQLMEKTKAKYF---SAFITYLQMKEMVNSDIPQSVLVNVCETLQSEVALF 1250  
DB 1813 SSVEKGOQLQOLLQRFYGHNNHHPPLDRQOSTIGE-----VIEAVKKEETW 1862  
QY 1251 DQTHSLALGSATEDKDSME--TDDCSRSHRDQDVCYVGLHLAKELCEVEDGDSWL 1308  
DB 1863 SQRDELVAIQCKUNDASDLVDIIMEPSRE--IFKAIHLARALL--HEGNDKV 1915  
QY 1309 QVTRRLPIPLTLTLEVSLSRMCQ-NLHFTBATLHLITLARTQO----- 1352  
DB 1916 QHS-----FYRMKQKQDIH--EPFKALITRIQTQNRKLKSDMWSCKSP 1959  
QY 1353 -----GATVAGA--GITQICLPLIS--VYQSTNGTAQTPSAS-----RKS 1391  
DB 1960 KVSSTVLTPLIDAGDTGFNGALFEVPOQVHPSPISEWSQSLNDSHSIPDLAPYQDEKS 2019  
QY 1392 LDA-PSWPGYVRLSMSIMEQLTKLR-----YNFLPEALDP-- 1426  
DB 2020 TDLALP-----EVALVEPILAVTLQCLCENHNSLQNFIRKQSDRTNHLVSTLSFLD 2072  
QY 1427 -----VGHQ-----ERTLQ-----CLNAVRTVQSLACEPADHTVGILO 1462  
DB 2073 TVCGSTKSGIVGEIGEHNFSLITQTLATLTFCQCPCHENQNTAMQENGLNIITSLV 2132  
QY 1463 LSNPMKEMPHLQVMDIQVNLGYLQACTSLHSKQMLQHYLQ--KNGDGLPSAAQR 1521  
DB 2133 LNEIKPLADDM-ELALEISQASKLALAMESRHDENANRVLRNANNSGGKQVHA 2191  
QY 1522 VQRPASASAPSSSKOPAD-----TEASEQALHTVQYGLKLSKTLALRH 1571  
DB 2192 IKQAVENTNNHMLKSIISDLFRQAEDDLKKSQPIYVTVTLPEINVDASIGVSI-H 2250  
QY 1572 FTFDVCOILIDQ-----SLDLAEVNF-LPALSFTPTFDSEVAPSGTILATVNVALN 1623  
DB 2251 TEKNISSSLDDKFNDDDIPSDVPREVGNHVIYILAHQALAHDEGE-----I 2296  
QY 1624 MGLSLDKKKEPLQAVGLSTQAGT-----RTLSKSLMTMENCFTL 1665  
DB 2297 WLDGDEKQDLTRE-ALNTYKERTAOIEIVRDRITLERVVPINDICSTL 2346

RESULT 23

A35774  
 Kinase-related protein sevenless - fruit fly (*Drosophila virilis*)  
 C:Species: *Drosophila virilis*  
 C>Date: 05-Oct-1990 #sequence\_revision 05-Oct-1990 #text\_change 09-Jul-2004  
 C:Accession: A35774  
 R:Michael, W.M.; Bowtell, D.D.L.; Rubin, G.M.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 5351-5353, 1990  
 A>Title: Comparison of the sevenless gene of *Drosophila virilis* and *Drosophila melanogaster*  
 A:Reference number: A35774; MUID:90313110; PMID:2115169  
 A:Accession: A35774  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2594 <MIC>  
 A:Cross-references: UNIPROT:P20806; GB:M34543  
 C:Genetics:  
 A:Gene: FlyBase: Dvir/rev  
 A:Cross-references: FlyBase: FBgn0013140  
 C:Superfamily: sevenless; fibronectin type III repeat homology; LDL receptor YWTD-contain  
 C:Keywords: ATP; autophosphorylation; phosphoprotein; transmembrane protein  
 F:9-26/Region: glutamine-rich  
 F:465-555/Domain: fibronectin type III repeat homology <FN3>  
 F:2232-2499/Domain: protein kinase homology <KIN>  
 F:2230-2238/Region: protein kinase ATP-binding motif

Query Match 1.6%; Score 144.5; DB 2; Length 2594;  
 Best Local Similarity 19.9%; Pred. No. 1.1;  
 Matches 335; Conservative 197; Mismatches 590; Indels 561; Gaps 79;

QY 328 LAVALHNTLPBETSVVRKIGTALQANVQYLTLRLSLASGNDCTTSTACMCVYG 387  
 DB 943 LRMAVQSGSL-----YTNELGGQLQPLPV-----QLASSP-----G 974  
 QY 388 LLSFVLRLSLHTLGNQDIIIDTACEVLADPSLP-----ELFW-----425  
 DB 975 PLALVNVSAVYVSGRGQSL--HCNVLLRQPSCTBERLEHGAVALVDMRGGLLHTDL 1031  
 QY 426 -----GTEPTSG-----LGIILDSVCNPFHL-LSPLLRLALVSGSKTAKV 468  
 DB 1032 ARDCVQRLDPFSGERELLPFGARHLALDSAG--HLVYSSSHLARRSLSLST---- 1084  
 QY 469 YSLPDKRSFYNELKHKPHVISHEDGTLWRQTPKLLYLGGQTN--LRIQGTIVGV 525  
 DB 1085 -----HQPELEYHVNG-----LAQISGFCIDLQRHLYWL 1116  
 QY 526 MLDRAVLVMEYSYSWTLFTCEIEMLLNVSTADVIQCHQKRPILDLVHKVISTDLS 585  
 DB 1117 VAGNSA-----LHLRYTALSAGSSQAVP-----1140  
 QY 586 IADCLLPITSRIYMLQRLTTVISPVDVIASCVCNCLTVLAA-----RNPA 631  
 DB 1141 -----LQLLTTL--PAADALPHTLQHLAPLGLMLLAADGRANHLRLAA 1183  
 QY 632 KWTDLRHNGFLP-FVLAHPVSSLSQMSARG-----MNAAGYCN 669  
 DB 1184 QLETTIDTMRKLRLEGVLEPLSAVQLERSAGPPRRPDEGVRLAVPPDSVHIDEGHNH 1243  
 QY 670 LLMNSEPQCG-----EYG-----VTIAFL--RLITTVYKGLQSGOSQGLVY 709  
 DB 1244 DFRVQRQPAASGNGHSVCYKLLHEHGERLITLLETPFARITQQLAARPLG----- 1294  
 QY 710 CVMFVLKEMLPSTHKRYNSHGVREQIC-----LLELIHALINLCHET-- 754  
 DB 1295 -----RISTPTPTARRAGS--TRVQDLDPVAAPTORRLRKFVERQAAFLQLAPNVSAAL 1348  
 QY 755 ---DLHSH--TSLQFLICSLAYTEAGOTVININGIDVTID-VWMAAQRPSDGAEEQ 808  
 DB 1349 LRNDVBEHAGSOSLQYRISC-----WRGSELHSELINOSTLEAREVHLQPEE--TYR 1400  
 QY 809 GQGQLIKTVKLAFTYNNVIRLKPSNVVSPLEQALSQHAGNNLIATV-----AKY 862  
 DB 1401 FQVQAVAAATGLAAGATSHALHVSPEVQSPRLLYNAAEH-----IGSLDLDTGHRKQ 1453  
 QY 863 IYHKHPALRLAIOQLKRLATVAPMSVYACLGNDAAALRDAFLTRLQSKIEDMRKIKMI 922

DB 1454 LVTTASPEVHLLVVLQSGRLIMV-----NEHVELSHVPGKAPALANRAREVLA 1503  
 QY 923 LEFLVAVETQPOLIELFLNLEVKDSDGSKESPLGWSCLAAVLELIDSGQDRYWCPP 982  
 DB 1504 LVTDWVQ-----RIYVAMELDAADG--GGVIVISLDCRDGNIL-----QGERLMSTP 1549  
 QY 983 --LHRAALAFLLA---LM-QDRRDSAMLVLRK-----PKWENUTSPFGLTSPPE 1030  
 DB 1550 RGQLLDLVALPHARQLVWLQHDLDNRNATLQGRSLANGSALTFEGVTLPLMWLF----E 1605  
 QY 1031 TSEPSILETCALIMKICILEIYVVVSGSDOS-----LKDTLKKSIFEKRFAY--WS 1080  
 DB 1606 GSGEPLEAFETINLVHDIGRLCVHVAHQCTSSALRYAQMLINDITQIADDPGLYALRN 1665  
 QY 1081 GYVKSIAVHVAETEGSSCTSLLEYQ---MLVSAW-----RMLLIATTHADIMHLD 1129  
 DB 1666 GSVRAGRRRRQOLE-----FILELQDEVRLRAYVQAVPSRRCLLPTTAALSTPS 1720  
 QY 1130 SVVRQLDFLDVLDGTAL-----LVPAVNCRLRGSMKCTLLILRLQMKRELSVDEIL 1185  
 DB 1721 SCEETQCSLQ-LPALSADPCPLPGLVQLWLSSS-----RSAQLELSLHSA 1771  
 QY 1186 GPLTEILEGVLQADQOLMEKTKAYPSAFITVYLQMKMKVSDIPQY-----1231  
 DB 1772 GLTINISQ--LQPYQAV--ELRAQVGSYYQQQLGQEPDLPVLTHTAATPSARPNSFG 1827  
 QY 1232 -----SOLVANCETLOEVIYALFDQTRSLALGATEDKSDMETDCCSRSHRDQDV 1286  
 DB 1828 RALSPSELELSWLPALPELSASAVYVTLHWOLQ---EDTEQSQOPAQOQGVETAGVQ 1883  
 QY 1287 CVLGLHAKELCGVEDDQSWLQ-----VVR 1312  
 DB 1884 RLTLGLQPAR-LIYQ-----WLDAAATPSKINSSGMLLRSTAPRLPDLIELNAYGMTL 1936  
 QY 1313 RLPIPLTLTTEVLSR-MKONLHTEATNHLTLTARQOGATAVAGAGITQSICLPYL 1371  
 DB 1937 AMRGTPDALSSLTLECGSLREGQLQFNVAGNHTQMLRAPQPK-----TRYSC--RL 1985  
 QY 1372 SVYQSLTNGTA--QTPSARSKSL-DAPSWPG-----VYRLSMS-LMEQLKTLN 1417  
 DB 1986 ALAAYAATPGAPYFGSHSEYETLGDAPSAPGRROLEHIGELFRVSWTALDNGSPILLY 2045  
 QY 1418 NFLPEALDFVGNQEBRTQLCLANAVRTVOSLACSEADHTVGFILQDSNMKEWHFHLPL 1477  
 DB 2046 NL--EHL-----QARRTRRRRRRRRETTLSLIPMAEPLVIBDQWLDPCNT--TELSCI 2095  
 QY 1478 MKDIQVNLGVLQOACTSLHSHRKMLOHYLQNK--GDGLPSAVAQRYORP--PSAAGA 1531  
 DB 2096 VRE-----LHTRRLLEFRVANRRPHGWPYSBDSERIAPFVSPFKRGS 2140  
 QY 1532 APSSSQAPADTBASQOALHTVQYGLKILSKTLAALHNPFDVQ-----IL 1580  
 DB 2141 LVLAIIAPAIYSSCV-----LALVLVRKLOKRRHRAKKLLQDSRPSIWSNLSAL 2190  
 QY 1581 LDQSLDLAEYNLFAISFTPTPFDEVAAPSFTLLATV--VALNML-----G 1626  
 DB 2191 QTOQQLAARSKTFMSLS-----DADIA-----LRLQIMWNLTLIRFLGSGAFGVYGG 2241  
 QY 1627 ELDKKKEPLTQAVGLSTQAEGRITLKSILM-----FTMENCFYLL-----ISQAM 1671  
 DB 2242 QLOAEDEAOPQRVAALKSIRKAGSEFAELLQEAQMSNFGENIVCLIGCDTDSISLIM 2301  
 QY 1672 RYLRDPAYVRKQMKQKELSESLTILSSLSRYRRRQPS--PATGVLPSPQSGTSL 1730  
 DB 2302 EHM-----EAGDLIS--YLRAPSSQELSKLOLELISMCLD 2338  
 QY 1731 KAS 1733  
 DB 2339 VAN 2341

RESULT 24

A40743  
 IPI receptor, XIPR - African clawed frog  
 C/Species: Xenopus laevis (African clawed frog)  
 C/Date: 03-Mar-1994 #Sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
 C/Accession: A40743  
 R/Name, S.: Muto, A.; Aruga, J.; Nakagawa, T.; Michikawa, T.; Furutachi, T.; Nakade, S.;  
 Cell 73, 555-570, 1993  
 A/Title: The Xenopus IPI receptor: structure, function, and localization in oocytes and  
 A/Reference number: A40743; MIMD:93258819; PMID:8387895  
 A/Accession: A40743  
 A/Status: preliminary; not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 1-2693 <KUM>  
 A/cross-references: UNIPROT:Q91908; GB:D14400; NID:g464219; PDB:BA03304.1; PID:g464220  
 A/Experimental source: oocytes and eggs  
 A/Note: sequence extracted from NCBI backbone (NCBI:131713)  
 C/Superfamily: inositol-trisphosphate receptor

Query Match 1.64; Score 144; DB 2; Length 2693;  
 Best Local Similarity 17.54; Pred. No. 1.2; 576; Indels 776; Gaps 95;  
 Matches 345; Conservative 270; Mismatches 76; Indels 776; Gaps 95;

QY 2 IRKSTKTSVLSFCRSHETLILGSALELSQIEALNKHWRRLLEGLSYKPPSS 61  
 DB 595 ITAAIDTFVSLVRKRE-----PRFLDYLSDLCVSMNK-----SIFVT 633  
 QY 62 AEKV-----KANKDVASPLKELGRISK-----LGIDEGSVOLLOCTIQ 102  
 DB 634 QELICRAVLPANADI---LIEKTLVSRFEFEGVAGSESSLEAGEDER-VWLFWRKSN 689  
 QY 103 EDVGRTRDSVKTVDROSOALILKADYY-YE-----ERTC----- 139  
 DB 690 KEIRSR--SVRELAQDKKEQEDQVLYTRQALNPAWKCLDROLYANETSGOLDVD 747  
 QY 140 -ILRCVLLHLYTPDERHPYRVEYAD-----VDLKEKELVS--KYRQOFELYTEA 189  
 DB 748 LILRCM-----SDENLFPDLRASFCRLMLHMHVDRDQEQVTPVXYARLMSIPEIA 800  
 QY 190 -PWETHGNL--MTER--QVSRFPVOCLEB-----QSMLELIT-----FLY 225  
 DB 801 IDVDSSGSTRDIDIKERFAQTMFEVEYRLVDVGQRPADKEKNKLTFFVNLARMLIY 860  
 QY 226 YAYFEMASPDLLVLTKEF-----KEQFGSRQTRNLVDDETM 263  
 DB 861 FGYVNF--SDLLRLTKLLAILVDCMILANFPGSKLKGESKSGSNVMSHGVGELMT 918  
 QY 264 PFVDRIQYF--SALILVEG-----MDIESLHKCALDRR--ELHOFADGLICQDMDC 312  
 DB 919 QVLLRGGGFLPMTPMARPEGTIKAQREPEKEDLVMDTKLIEILQFILNVLDYRISC 978  
 QY 313 LMLTPGDIPIHAPVLLAMALLRLHTLANPEETSSVVRKIGTALQLNV-----FOYLTRLL 366  
 DB 979 LILCIF-----KSEFDESNAGSVEGSTEALTVVPGTLDPEHIEQA 1018  
 QY 367 QSLASGNDCTSTPACMCVGLSFLVLSLEHTLIGQODIDPACGVLDAPSLPELFWG 426  
 DB 1019 EGIFFGSEF-----NTPLDLDGGR----- 1039  
 QY 427 TEPTSGILILDSVCGMPLHLSPLQLL-----RALVSGK--STAKK 467  
 DB 1040 -----TLFRLVLLHTMHQYRPLVSGALHLRLFRHFSQROVLAQFAQVQLVLTSGQVDNVKQ 1095  
 QY 468 VYGFPLDMSYF--NELYKAK--PHUVISHEDGTLRRQTP----- 503  
 DB 1096 IKODLQRLSRIVESSELMVYKSGSEVETAAQAGAKGETPGAKKSESTSSNYRVVK 1155  
 QY 504 KLLYPLGGGONLRLPGSTGVGMVMDRAVYVRMEYSSTWTLFCSEIML-LHVSTADV 562  
 DB 1156 EILRL--SKLCVQENTTGRNRKQOORILRNKAIS-----VLELLQIPREKTEDT 1206  
 QY 563 ICHQGVKPIIDLNVKVIIST-----DLSIADCLLPITSRIYML--QR 603  
 DB 1207 -----RKQELIMKLAHEFLQNFCAQNGQNALHNGHINVLTPRGILAEVLTQHIFMNNFOL 1261

QY 604 LTTVISPVDVIVASVNCCLTVLAARNPAKWTDLRHGFLPFY-----AHPVSSLSOM 656  
 DB 1262 CSSEINRWOFHAFHLE-----TIGRNVQYIKFLOTTIVKAGRIYKKQODI 1307  
 QY 657 ISAGGNMAG-----GYGNL--LNNSEPOGE-----YGVYIAFLRLITTLVK 697  
 DB 1308 VMLELVNSEDVAVFYNDRASFTVLQVMRSEBERDENPLMYHILHVLAVCTEGKN 1367  
 QY 698 QLGSTOSQGLVP-----CMFV-----LKMPLPSYHKRY 727  
 DB 1368 VYTEIKNSLILPDIIVRVVTHEDCVPEVXIAIYNFLNHCYVDTEVEMKEIYTSNMWK- 1426  
 QY 728 NSGVNRQIGCLLEILHAILNCHETDHSHTPSLQFCISLAVTEBGQVIMNGI 787  
 DB 1427 -----LENFVLVDICRVGNST--DRKLD-----YLERVYTELVMSIVN--- 1465  
 QY 788 GVDITDMVMAQPRSDGAEQGGOLLIKTVKLAFSVYNNVIRLKPSPNVSPLEQALSQ 847  
 DB 1466 -----TFSSPSFDQS-----TTLQASKLAV-----VHQTRQP--VFVQLQALFR 1504  
 QY 848 HGAHGNLLAVLAKYTH-----KHDPALPRLAIQLKRLA--TVA--PMSVYACLN-- 896  
 DB 1505 -----YHGMWLLPSQKASYESCIRVLSDVAKSRITAI PVDLDSQVNNLP 1549  
 QY 897 -----DAAITRDAFLT-----RLQ--SKIEPMRIKVMLEF 925  
 DB 1550 LKSHNIVQKTAMGWRMSVRNARBDVLTASRDYRNIETRLQDIVSLEL-RLEPVLQAE 1608  
 QY 926 LTVAVET--QPGLEIFLNLVNDGSDGSEFSLGMSG--LNAVLELDSQQODRYWC 980  
 DB 1609 LSVLVVDVLRPEL-----PEPNDARKKCSGGQFCIGLIHNTKLLLENEEK----- 1656  
 QY 981 PLLHRAALAFHALMODR--RDSAMVLRKPKFW-----ENLTS--PLFPT 1024  
 DB 1657 --LCIKVLTQRLREMAKDRFGDKGEALRYLVNRYGNIRPAGRRSLTSFGNGPISGG 1714  
 QY 1025 LSPSS-----ETSEPSILETCALIMKIIICLEIYVYVKSLSQSLDQTLKKPSIEK 1074  
 DB 1715 SSGKSGGSGTSGSLSRSEMSLSD-----VQCHLDKQGSGLDVI--DLINMATSDR 1763  
 QY 1075 RFAYWGSYVKSIAVNAETEGSSCTSLSEYQMLVSAMRMILLATTHADIMHLDTSVVR 1134  
 DB 1764 VFH-----ESILALALEGGNT-----IQHSFCRLTEDEKSE 1798  
 QY 1135 QLPLVDLDTKAL--LVPAVNCYLRGSMKCTLLILRLQWRKELGSDVDEILGPTEI 1191  
 DB 1799 KFFKVFEDRKVAQOEIKATVNTSDLSKED--EEOSERETPIHQVAREPSGOI 1853  
 QY 1192 LBGVLADQQLMEKTKA--KVF-----SAFTV 1217  
 DB 1854 TE--EAKQELIRASVATKKAAYSFPREADPDHFSLSEGVMAVAERKDELEMSAVITI 1910  
 QY 1218 LQ-----MKEMKVSIDIPOYSQVLN-----VCETIQ----- 1243  
 DB 1911 MOPILRLQQLCENHRDLQNFRCQNNKNVNLVCTTLOPDLDCIGSTTGIGLGLYI 1970  
 QY 1244 -BEVIALFDQTRSLALGATBEDKDSMETDCSRSHRRDQDVCVL-----GLHLAKEL 1297  
 DB 1971 NEKVALINQLTLESL-----TEYCOGCHENQ--CIATHESNGIDITALL 2014  
 QY 1298 CEVDEBDSMLQVTRRLPILPTLLTTLVESLRKQONHFEATLHLLTLARQOGATAV 1357  
 DB 2015 TLND-----INPLARKRDVLLELKN--ASKLLALMESHH----- 2049  
 QY 1358 AGAIGTQSLPLSLVYQSLSTNGTAOTPSASRSLDAPSWGVYRLSMSIMEQLKTLRY 1417  
 DB 2050 -----DSENAER-----ILY 2059  
 QY 1418 NPLPEALDFVGVHOERTLQCLNAVTVQSLACLSEADHTVGLTLQLSNFKMNFHLPL 1477  
 DB 2060 NMRPEKL--VEVIRKAYLQ-----GEVEFEDGESG-----EDYASPR- 2095

```

QY 1478 MRDIQVNLGYLCOACTSLHSRKMLOHYLQNKNGDGLPSAIVORVORPPSAASAAPSSSK 1537
DB 2096 -----NVG-----NHVYLAHQALAHNNE-----LQHMURP-----GV 2123
QY 1538 QPAADTEASEQOALHTVQYGLKILSKTALRHFTPDVCCILDDOS 1584
DB 2124 QTBGDEALBFYAKHTAQIEIVR-LDRTMQIVFPVPSIGEFLLKES 2169

RESULT 25
S45340
FKBP-rapamycin-associated protein (FRAP) - human
C/Species: Homo sapiens (man)
C/Date: 10-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C/Accession: S45340
R/Brown, E.U.; Alders, M.W.; Shin, T.B.; Ichikawa, K.; Keith, C.T.; Lane, W.S.; Schreibe
Nature 369, 756-758, 1994
A/Title: A mammalian protein targeted by G1-arresting rapamycin-receptor complex.
A/Reference number: S45340; MUID:94277209; PMID:8008069
A/Accession: S45340
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-2549 <BRO>
A/Cross-references: UNIPROT:P42345; GB:L34075; NID:G508481; PIDN:AAA58486.1; PID:G508482
C/Genetics:
A/Gene: GDB:FRAP1, FRAP; RAFT1
A/Cross-references: GDB:597698; OMIM:601231
A/Map position: 1p36-1p36
C/Superfamily: yeast TOR2 protein

Query Match 1.6%; Score 143; DB 2; Length 2549;
Best Local Similarity 17.6%; Pred. No. 1.3;
Matches 262; Conservative 246; Mismatches 559; Indels 420; Gaps 65;

QY 237 LVTTKFKEGFGSGRQNRHLVDETMDP-----VDRIGFSLIIVEGMDIE 284
DB 217 LITLTQREPKEMOKPQWYRHTFEAEKGFDETLAKENMRDRRI-HGALLIINELVRIS 275
QY 285 SLHKCALDDBRELHOFADGLI-----CODMDCMLTFGDI PHHAPVLLAWALLRHITLNP 340
DB 276 SMGGERL--REHEBEITQOOLVHDYCKD---LMGGTFPRHITPTTSF---QAVQPO 325
QY 341 ETSSVVRKIGGTALQNLVFOYLRLLQSLASGNDCTSTACMCVYGLSFVLTSLHT 400
DB 326 OSNALVGLGYSSHQ-----GLMGF----- 345
QY 401 LGNOQDIIIDACVLDPSLPFLFWGTEPTSGIILDSVC-----GMFPHLS----- 449
DB 346 -----GTPSPAKSTLVESRCRDLMEKEFDQVCOWVLKC 380
QY 450 -----PLQLLRALVSGSKTAKVSYFLDKNSFYNELYKHKPHDIVISHEGTLWR 499
DB 381 RNSKNSLIQWTILNLPRLAAPPSPA-----FTD-TOYLODTMNHVLSCKKKEKRTAAR 434
QY 500 KQTPKLLIYFGGTNLRIPQGTVGQVMD--DRAVLYRMEYSYS-----SWTLFTCEIE 551
DB 435 QALGLLSVAVRSEFKVLPFR-----VLDTIRALPDKDPAHKQKAMQVDAIVFTC-IS 487
QY 552 MLHVVSTADVIOHCQRKPIIDLVHKVISTDSLADCLLPITSRYMLQRLTYYISPP 611
DB 488 MLARAPGPG--IO-----QDIKELLEPMLAVGLSPA-----LTAVLYDLSRQIPOL--K 532
QY 612 VDVIVASCVNCLTVLAARNPAKVTDLRHGTFLPFVHAPVSLSQMISAEGMAGGYNLL 671
DB 533 KOIQDGLKMLSLVLMHKP-----LRHPCMPKGLAHQLASPLTLTPREASDVG----- 580
QY 672 MNSEQPGQEGVGTIAFLRLITTLVKQO-----LSTQSGVLVPCVMFVLKEM 718
DB 581 -----SITLALRTLGSEFEGHSLTQFVRHCADHPLNEHKEIRMEARTRSL 630
QY 719 LPSYHKRVNASHVRQOIGLLIELHAL-----NLCHEIDLASSHT 761
DB 631 TBSIHLSGHAHVVSQTAQVADVADVLSKLVLGVTDPDDIRYCVLASLDERFDALHAQ 690

```

```

QY 762 PSIQFLCISLATEAGQYVINMGIGVDTIDVMAAOPR-----SDGAEQ 808
DB 691 ENLOAL-----FVALNDQVEIREIRLAICTVGLSSMNPAPFAPFLKMLIQITTELEHS 744
QY 809 GOGOLLIKTVKLAFTVNNVIRLKPSSNVSPLEAL-----SQHAGNNILAVL 859
DB 745 GIGRIKQSKRMGLHVSNAFRLIR--YMEFLIKLILKADPDPPNGVYNNVALTI 802
QY 860 AKYIYKHDPALPRIALIOLKRLATVAPMSVYACIGNDAALIRD-----AFLTRLOSKIE 914
DB 803 GE-----LAQVSGEMRKVNDLFIIMDMLODSSLAKRQVALW 842
QY 915 DMKIKMIEFLVAVETQPGLEFLPLNEKVDGSGSKSEFSGMMSCLHAYLELDSQ 974
DB 843 TLGQALVATGYVPEPRKYPTLLEVLINLTKTEQNGTREAIRVGLGAL----- 894
QY 975 QDRYWCPRPLLRALIAFLALMODRDSAMLVLRTRKFWENLTSPFGTSPSPSETSEP 1034
DB 895 -----DPTKHKVNTGMID---QSRDASV-----SLSSKSGDS 926
QY 1035 SILETCALIMKI--ICLETIYVVGSLDQSLKDTLKKFSIEKRFAYMSGYKSLAVHVAE 1092
DB 927 SDVSTSEMLVNMGNLPLDEFYPAVSWV-----ALMRIFRDQSLSHHTMVQATFIFK 980
QY 1093 TESSCTSLILEQVMLVASMRMLIATTTADIMHLTDSVVRQLPLDV----- 1140
DB 981 SLGKCKVQFLPOVW-----PTFLNVIYRCDGAIREFLFQOLMLVSVFKSHIR 1028
QY 1141 ---LDTGKALLLVASVNCRLGSMKCTLLIILRLQMKRELSGVDLIGTLTLEBVLQA 1198
DB 1029 PVMDEIVTLRREHWNT-----SIGSTIIL-LIEQIVVALGGEFKY--LPQIIPMLAV 1081
QY 1199 DQQLMEKTKAVPSAFITVLQMKEMKVSIDIPOYSQVLNVNCTLOEVALPDQTRHSI- 1257
DB 1082 ---FMHNSPGRIVS--IKLIAIQLFGANLIDYHLILL-----PRIYLFDAPEAPLP 1130
QY 1258 ---ALGSATBDKDSME--TDQSRSHRQDRGVCTGLHLKKEICEVEDDQSWI--QV 1310
DB 1131 SKKALETVDRLTESLDFTVYASRIIH-----PIYRITDQSEBELSTAMDLTSSIVFDL 1184
QY 1311 TRRLPLFTLITLLEVSLEMKONLH-----FT-----EATLHLITLARTOQAT 1355
DB 1185 GKKYQIFIPMNKVLVRRHINQRYDVLICRIYKGYTLADDEEDPLIYQHRMLRSQGA 1244
QY 1356 AVAAGITQSICLPILSVQSLNGTAQTPSASRSLDAPSW--PGYTRLSMSLM--EQIL 1413
DB 1245 LASGPVETG---PMKQLHVTINLOKAWGAARVSKD--DWLEWLRRLSLLELDSSSP 1298
QY 1414 TLR-----YNFLPEALD-----FVGV-----HOERTQCLNAVATVQSLACLEEA 1453
DB 1299 SLRSCMALAQATN--PMARDLPMNAFVSCWSELNEDQDELRISLTALTSDIA---EV 1353
QY 1454 DHTVGFILQLSNPKMEWHEHLPLQMRDIO--VNIG---YLQACTSLHSRKMLOHYLON 1508
DB 1354 TOT---LNLALFEM--EHSKGPPLRPDDNGIVILGSRKAKCAVAKALHYKEI--EFGG 1407
QY 1509 KNGDGLPSAVA--QVQRPSPAASAAPSSSKQ--PAADTEASEQOALH 1552
DB 1408 PTPALIESLISINNKLOPEEAAAGVLEYAMKHFGELEIQTATYERKIH 1454

RESULT 26
D88450
protein P21H1.2 [imported] - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C/Accession: D88450
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A/Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A/Reference number: A75000; MUID:99069613; PMID:9851916
A/Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_eleg

```

A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
 A:Accession: D88450  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Releases: 1-2700 <STO>  
 A:Cross-references: GB:chr\_III; PIDN:AA19433.1; PID:g506795; GSPDB:GN00021; CESP:F21H11  
 C:Gene: F21H11.2  
 A:Map position: 3

Query Match 1.6%; Score 142.5; DB 2; Length 2700;  
 Best Local Similarity 17.8%; Pred. No. 1.6;  
 Matches 334; Conservative 261; Mismatches 661; Indels 619; Gaps 92;

145 LHLLT-----YFOEBHRYREYVADCYDKLEKEVSKRQGFEL- 184  
 22 LHIYEEBPLEKMLNKTLORGGDLYFDMLCHTLHGLSEICLPPLIKVLVEMYEKDESLC 81  
 185 -----YKTEAPWETHGNLWTERQVSRMFVQCLREQSMLEIIFLYAYAFEMAPSDLL- 237  
 82 LSLMSPITATPELRKLKHLKGLAVN-----YLFCL-----VLIEI--LPQVEHLLPQCDPLI 130  
 238 --VLTQKPE-----QGFGRQTNRLHVDETMDPFVDRIG--YES----- 273  
 131 KVLVEICFKNQVREPSAVGINKTNHLVVAETGEVGVLSSTYETTHIRHIFMTHITELK 190  
 274 -----ALLVEGDIESLHKCALDDBRELHQFQDGL-ICQDMCLMTF--GDI 320  
 191 KDVSOTAAQOIVAIIMSKPLRINSQVED-----FENGKLFDDGSLLEVKQDV 243  
 321 PHHAPVLLAMALHRTLHPEETSSVVRKIGGTAIOANFOYLTRLLOSLGASGDCTTS- 379  
 244 KHAWGLLVELL-----PVAQIKR-----ETNI--PALISLVQKLYTTNDMSKK 289  
 380 -----TACMCV-----YGLSPV--LTSLEHTLGNQODIIDTACEVLADPSLEL 423  
 290 OHKLAAPLITCLCVSQKHFFLANWQFLNSCLSHLKNKDPQVAVALE-----SLYRL 344  
 424 FW-----GTEPESLGIILDSVCG-MFP--HLISPL--LQILRALVSGSKTKKAYS 470  
 345 LMVMIRNADGNATRSRLDSICSLRPFKNRYLVPPDAPLANIFVKIHLISQKIDFA 404  
 471 FLD-----KMSFYNELYKGRPHDIVSHEDGTLMRQTPKLLYPPGGQNTLRIP 518  
 405 FKEIIFDLCCVNNRQGRSLYAEKRNNGVIRALMVAIDGLOQKDDPRAMKSKSPSAS-- 460  
 519 QGTGVQVMDRAVLYRWEYSYSWTLFTCEIEMLLHVVSTADVIQOCQRYKPIIDLVHK 578  
 461 -GTVHKT--KRYQVYTR-----PLTNEISKSIGIDQYPPQCRKADSIIR 502  
 579 VISIDLSIADCLPTTSRIYMLQRLTVISPPVDVIASCVNCLTVLAARNPAKWTDLR 638  
 503 LLDPIQIGPLMMSSIQNGKKEPDELISGDAKPKDLFRTCI-----AALPRLIPDMS 555  
 639 HTGFLPFAHNVSSISQMSABGNMAGVGNLMSSEQGEYGTIAFLILITTVKQ 698  
 556 HVDLIDL--LTRLTVHLEBELRMSGITLQTIIGEPDMREBOVFISHISLSIOHI-- 608  
 699 LGSTOSQGLVPCVM-FVLKEMLPYHKMR--YNSHGVEQIGCLILEIHAHLINLCHETD 755  
 609 -----YFFQIILDDSLRLMLQSLITWKTATIAKRE-----AKIILPSPECHSH 657  
 756 LHSHTPELQPLCISLAYTEAGQTVINIGIGVDTIMVMAOFRSPGABEQGQGLLI 815  
 658 VTSNSTHSTHISVSEPNVTVSASISLSSSSPHQTPS--LCSLPRESSLSHS-----I 708  
 816 KTVKLAFSVTNNVRLKRPSPNVSPLEQALSHGHGNLILAVLAK-----YIYKHDP 870  
 709 PTAMTSLSTTTTTHHSQAPSTSTPI-----GSSGNTLSVLHQMGLAIIVYCQTRS 761  
 871 LPR-LAIQLLKLATVAPMSVYACIGNDAAIRDAFLTRLQSK-IEDMRIKVMI---LEF 925  
 762 NPRAIAVILKEIQIILYDGLGIEIMDTVIDVLDQATPYVVKYIEHPIKERMSKMNDF 821

926 LTV-----AVETPGILEFLNLEVKDSDGSKESFGMSCL---HAVLEIDSOODR 977  
 822 SSVCDDKSTIETDNTLV-----NSDRGENEYFQNDPMGCALSGAIEIHLITK----- 868  
 978 YWCPRLHRAAIAELHLMODRDSAMLVLTQKPKFENLTPPLFGTLP--PSETSEPS 1035  
 869 --CP-----SAVA---AAM-----FILPSRLNA-VSGYVDPNPNQNESRSS 903  
 1036 ILETICALIMKICLETIYVVGKSLDQSLKDTLKFKFSIEKRPAYSGYVKSILAVVALEEG 1095  
 904 LL-----RGS-----XSKG 912  
 1096 SSCSTILEYQW---LVSAMRMLIATTHADIMHLTDSVVRQLPLVDLGTALLVLP 1151  
 913 TS--SILGEOLGQACALMOKYILMCA-----LAP 942  
 1152 ASVNCRLRGSMTCTLLILRQWRRELSYDEILGPTLEIGVLQADQOLMEKTKAVF 1211  
 943 APYNTLS-----QRSFPTSMQDPV-----DVF 965  
 1212 SAFITVLOKMKVSDIPQISQVLANCETIQOEVIALFDQTRSLALG----- 1260  
 966 RSVASLSRSRTPVPN--SISQLSIKVCMILRMENLT--DIRDSVVLGVGSINPLAFDM 1020  
 1261 -----SATEDKDSMETDSCRSRHRQDR-----GVCGLGHLAKELCEVED 1303  
 1021 MLEELKNGILREATEK--AETN-----LRRKKKDLRLQITIVIEVAIRGLLHSA 1074  
 1304 GDS-WLQVTRRLPLIPLTLTLE-----VSLRMKONLHFEATLHLLTLARTQOGATA 1356  
 1075 GSSDFILHPHVVDIDSMRVLBSHDHRLIVYTRKLHFAK-LIHLLID----- 1123  
 1357 VAGAGTOSICLPLLS-----VYOLSTNGTAQTPSASRSKSIDAPSMGVYRIS 1404  
 1124 -----STPHLSHHTLPSDERRHKLFTLFINMCSRAIAADRKFKDKEVGSYVEQS 1173  
 1405 MSLMEOLL-----KTL-----RYNPLBEALDPVGVHERTLOCLANAVTQSLACL 1450  
 1174 VLANSRLILCCGPITREPASKSIGEDGYLIGMLEKLI--ISTNPTWQA----- 1216  
 1451 EADHTVGFILQI--SNFMKEW-----HFHLPLQI-----MRDIO--VNLGY 1487  
 1217 -EVEBMLAMMELMESGVLDMLMOCYTOPPYACGCFRALVRFSSRDPCEFSLSFV 1275  
 1488 LQOACTSL-----LHSRKM-LQHYLONKXGDGLPSAVQRYRPPSASAAASSSKOP 1539  
 1276 LCOGMLAVNSVTDICALHMBILIRKQPLETSN-----IHATSPAQCVTPIVQWRN 1324  
 1540 AADTEASEQALHTVOGLKILSKTLAALHFTPDVCOQLLDGSDLAENPLFALSFT 1599  
 1325 QADVLNTINYNGH-----ILPIEGH--DVCTRLAN-----STPHLTVTIFS 1363  
 1600 TPTF--DSEVAPSEGTLLATVNALMIGELDKK-----KEPLTQ----- 1637  
 1364 EVSYRLTESGNSNLSQALLALQPMISNLELVQDVVEEAAGPRMGSEBETQVLNML 1423  
 1424 YLVTVLSDNDEKELAESLPLFTYRIC--VMSQVVGTRISSILLEWL--STVHDSKXT 1479  
 1638 --AVGLSTQAGTRTLKSLMFTWNCIFYLLISQA-----MRYLRDPAVHPRDK- 1684  
 1480 LDRSEIIPPYRWKDE-NSEKRVANGSIAKDSLEEBPTKEGVRLLPMPAYGHHYSPLOSQF 1538  
 1733 SPSEQPLQVQAF 1747  
 1539 LP-----PYVQPVQFF 1549

RESULT 27  
 A54837  
 rapamyclin/FKBP12 target 1 - rat







[illegible]

```

Db      : 953 PRSFTTSLCIRMDSEKWDV---TTIGSTASWMLSTICSSHPNSFLVV-MEGNGLEI 1008
Qy      : 1017 LTPSPFGLTSPSETSEPSILETCALIMKIICLEIIYVVKGSLDOSLKDYLTKESIERKF 1076
Db      : 1009 IAEULQRKXSTQOENSSDESKWTAMSTLAWSQBPKVSSPATENILQTLAPFMQSGM 1068
Qy      : 1077 --AYMSGVYSIAVAHVAETEGSCSTSLLEYOMLVANEMLLIATTHADIMHLTDSVVR 1134
Db      : 1069 IDGFYTAQV--LAALVRHKNDKTISEIMNIDIVETT--INLVCEESDTRSLCALABEL 1123
Qy      : 1135 QLFPLVDLTGYKALLVPASVNCRLGSN--KCTILLI-LERQMRELGVSUDEILCPLEI 1191
Db      : 1124 SLVONPYEATELEVIFENERV---RSGSFKKCIPLVLNLRKPYADKVGGIPVARLLRRI 1180
Qy      : 1192 LEGVLQAQQUMEX---TKAVPFAFITVLQMKEMKXSDIPQYSQLVINVCETLOEBVIA 1248
Db      : 1181 ----ANDDSLKLLIAEAGALDALAKYLSLSPODSTEI-----TVSELBS 1222
Qy      : 1249 LF--DOTRHSLALGATEDKDSMETDDCSRRHRDQRGVCLG-IHL-----AK 1295
Db      : 1223 LFRSEITRRHTAISSMQ-----LGIHLBASRTRYNAAR 1259
Qy      : 1296 ELCEV---DEDGSMLOVTRRLPILPULLTELVLSRMKNÜHLFEATLHLTLTARTQO 1352
Db      : 1260 VLCELFSESHIRDESLAWKALSPILEMNTLE-----SERVALTALVKLTMGINPRPD 1314
Qy      : 1353 GATAVAGGITQSICLPILSYQLSTNGTACTPPASRSKLDAWPMGVYRLSMSMEOLL 1412
Db      : 1315 ILTSLEGN-----PLDNITYIKI-----LSLDSS-----SLESKTSAA 1345
Qy      : 1413 KTLARNFLPEALDPFGVQHERTLOC-----LNARVTQSLACEADHTVGFIOLSNFM 1467
Db      : 1346 RICREFLTFNE-----GLRTSTSAACCVISLISIRTGSTV-TERGMALDRDLDIRFV 1399
Qy      : 1468 KEWHFHLEPOLMRDIQVNLGYLCOAQTSLHSRKMLCHYLQNKNG-----DGLP--- 1515
Db      : 1460 EVAEBEH-----DCVNLFYGVVASNNVLISEBAISLCITRWAKONTPRKM 1442
Qy      : 1516 -----SAVAORVOPRPASAABASSKHOPADYE--ASEQOALHTVOYGILKITUSK 1564
Db      : 1443 DLIMKGIIEKSIOQSKSPSSLCSVIADLFRVLTVNGVIARSODAIKWQPLLLILRQ 1502
Qy      : 1565 -----TLAARHTPTDVQCJILDOSLDLAAYNFLPALSTFTTPFDSEVAFSGTILATV 1618
Db      : 1503 DUDFOGQGGIOAQIANILIEKPMVLESILIASSTIMPL--IPLIESE---SIAYKAT- 1555
Qy      : 1619 NVALNMUGELDKKKE-----PIQVAGLSTQABGTFTLKSLIMTWME 1660
Db      : 1556 TILTSLTEMORPOBEITTKULIAPLVAVGIR-----VRNLOETALMGLE 1601

```

## RESULT 29

T50176  
 probable peptide synthetase [imported] - fission yeast (*Schizosaccharomyces pombe*)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jul-2004  
 C:Accession: T50176  
 R:Badcock, K.; Churcher, C.M.; Wood, V.; Barrall, B.G.; Rajandream, M.A.  
 submitted to the EMBL data Library, February 2000  
 A:Reference number: Z25044  
 A:Accession: T50176  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-4924 <BAD>  
 A:Cross-references: UNIPROT:Q9P7T1; EMBL:AL138654; PIRID:CAMB72227.1; GSPDB:GN00066; SPDB:  
 A:Experimental source: strain 972h(-); cosmid c23G3  
 C:Genetics:  
 A:Gene: SPDB:SPAC23G3.02c  
 A:Map position: 1  
 A:Introns: 2125/1  
 A:Keywords: carrier protein; phosphopantetheine; phosphoprotein  
 F:260-697/Domain: acetate-CoA ligase homology <ACLI>

F:2772-3226/Domain: acetate-CoA ligase homology <AC12>  
 F:4405-4474/Domain: acyl carrier protein homology <ACP>  
 F:2206,3288/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 1.6%; Score 141; DB 2; Length 4924;  
 Best Local Similarity 19.7%; Pred. No. 5.3;

Matches 203; Conservative 137; Mismatches 369; Indels 320; Gaps 49;

```

QY 752 HETDLHSHTPSL-----QFCICSLAYTEAGQVYINMGIGVDTIMV--MAA 798
DB 707 HSTDSRSSNQSLELVANLHETCEKRENSVVIS-----RFSIFDGLDSDLVYFVR 760
QY 799 QPRSDGAGQGQGLIKITVLAFTVNNVRLKPPS-----NVVSPLEQALSGHGA 850
DB 761 KIRSLGFPANPS--IVLTSKVVFKLASLLIKNEVOSKNFLNKCTPIQKGLYSEF 817
QY 861 HGNMLIAVLAKYIKH-----DPAFLPAIQKLKLAIVAPSVYACGNDAAIRD 903
DB 818 SNNG-----NLFFNHVVFKLAPSPKVKLAMEKL-----LDHTTILSNGFALDEN 862
QY 904 AFLFR--LQSKIEDMRKIMLEFLTVAVETOPGLIEFLNLEVQGS---DGSKEPSLG 958
DB 863 EGFTRFLEKKPPLYSYKNCLECIQKGFTRKE--FDEQFLNSGFLDAVYIDSNCVUST 921
QY 959 MWSCLAVLELIDSGQODRYWCPLLRRAIAFLHAWQDRDSAMLVLRTPKRWENTLT 1018
DB 922 VMH--HALYD-----GWSIDIIMQGLFHLH---DR-----LTIYQPED--- 957
QY 1019 SPLQGLSPSETPSESLTELICALIMKICIEIYVVGSLDQSLKDTLKFSLEKRPAY 1078
DB 958 -----YV-----QEL-ESLRNLANKNCISF 976
QY 1079 MSGYV-----KSLAVH---VAETEGSGCTSLLEYQ-----MLVAMRMLLIIA 1118
DB 977 MKKYLKDKFKESLSYQREKMGVLELSNISLSVENICOKIQTTPLSFLTANSTVL--- 1033
QY 1119 TTHADIMHLDTSVVRRLFDLVDLGTKALLIVPASVNCRLGSKCTLLILLRQMKREL 1178
DB 1034 -----SSYIKTN---DPLVGT-----VVGKGVNSL-----L 1056
QY 1179 GSVDEILGLTEILEGVLQADQOLMEKTKAK--VFSAPITVLQMKEMVSDIROYSOVLIN 1237
DB 1057 PNVIYVIGPCWQTPVRIKLDDELSYKQLCONLFRHSFVLKHSVMAISDE--QSELLVSN 1115
QY 1238 VCETL-----OEVIALPDQTRHSIALSASATEDKDSMETDCSRHRHQRDGV 1286
DB 1116 LFEISILYIQSGGPSVBSFISLHSTDH-----VEQPLRIEIKNGYKFK----- 1162
QY 1287 CVLGLHLAKELC--EVDDEGDSMLQVTRRLPIPLITLTLLEVSLMKONLHTEATLHLL 1345
DB 1163 --LTVGLSSELNNLNDPDK-----ILNFIIVNIESKIQTHASFVNTTIEHNHVE 1211
QY 1346 TLARTQCATVAVAGITOSICLPLLSVYOLSTNGTQTPSASKSISDAPSPGCVYLSM 1405
DB 1212 SKATL---GFSKREKLVRSCLSLKILGNTVLSAVFERLQNGDSDCLANRFSBLKSS 1268
QY 1406 SLMEQLIKTLRYNLFPEALDFVGNHVERTLOCLNAVTVOSLACLEBADHTVGLIQLSN 1465
DB 1269 GIGMLKIPMKSNPFIASL-----CELLVLPTELTLS---ADNRIETP--YEVSD 1311
QY 1466 FMKEMHHLPOLMDIQVNLGYLQOACTS--LHRSRKLQHYLONKRGDGLPSAVAOVQ 1523
DB 1312 IGHANL-----DIESFOYFPTPMQOALLASSEKNGVEYYNKK----- 1351
QY 1524 RPPSAAAPSSSKOPADTEASEQOALHTVOYGLIKILSKTLAAR--HTTP-----DV 1576
DB 1352 -----LFTCKSSQEBE---YLLFTLLANLPILRTCTFFVTRSKYYPY 1391
QY 1577 CQILLDSLDLAEXNPLFALSFTTPTPDSVAPSGFGLTLATVNVVALMMLGELDKKEPLT 1636
DB 1392 CQVLANE-----PNFYFOVLPHYKGSLSKYNLAELPL--LDSKVPPI-- 1431
QY 1637 QAVGLSTQAGSTRTLKSLMFTMENCYLLISQMRYLDRPAVHPRDKQMKQELSELS 1696

```

Db 1432 QIFFL--QGEN-----KNYVFCIHNVLY--DANAFOIIMDINHILKRNPKG--SQSML 1481

QY 1697 TLSSLSRY 1705

Db 1482 KFIISYLHRY 1490

# RESULT 30

H86502 excludase ABC subunit A [imported] - Chlamydia pneumoniae (strain J138)

C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

R/Shirai, M., Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shibata, T.; Ito

Nucleic Acids Res. 28, 2311-2314, 2000

A/Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A/Reference number: A86491; MUID:20330349; PMID:10871362

A/Accession: H86502

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1826 <STO>

A/Cross-References: UNIPROT:Q92985; GB:BA000008; NID:98978469; PIDN:BA98306.1; GSPDB:GN

A/Experimental source: strain J138

C/Genetics:

A/Gene: uvrA

C/Superfamily: Chlamydia trachomatis probable excludase ABC chain A; ATP-binding case

Query Match 1.6%; Score 140.5; DB 2; Length 1826;  
 Best Local Similarity 19.0%; Pred. No. 1.1;  
 Matches 365; Conservative 234; Mismatches 610; Indels 717; Gaps 95;

```

QY 223 FLVYAVEMAPSLV--LTKMKEGFSGSRQINRHLDVETMDPFVDIGYFALILVEG 280
DB 92 FSHYSHATVSTIELSHLALFTLEG---QARDPKYEVLDLYSKKYLSTIMELSEG 147
QY 281 MDIESLHKCALDDBRELHOPADGLICODWDCLMTFGDIPIHNAPYLAWALRHITLMEP 340
DB 148 VQSLIPLRLKIDIAIHVEAQQ-----FTKYRCNGTIIPIYSFTLSGI--PE 194
QY 341 ETSVSVRKIGTAL--QANVQYVLTLLDLSLGGNDCTTSTACMCVYGLSVYLTSL 398
DB 195 DCS--VDIVDTLIKENNIRARLVSLLFTALFEGEHC-----VLDBDEL 238
QY 399 HTLGNQODIDTACEVLADPSLELPFGTEPTSGLIILDSVC-----GMPPHLSPIIL 452
DB 239 MTSTIQQIDD---VTPYPLQQLF---SPHA-----LESRSLOCGSGIFISINDPLL 286
QY 453 QILRALVSGKSTAKKYVSPIDKMS--FYNELYKHKHVDVISHEDGTLWRQPKL--LYP 508
DB 287 -----IDENLSIKENCCSFAGNCSSYLHTIYQALA--DALNFULETPMKOLSP EIQNIF-- 339
QY 509 LGGQTLKRI PQGVQVM----- 526
DB 340 LRKKNMLVLPVRLFDOTLGRKNTLVYVWRGVLDIGDKVRYTTPKPSRYLSKGNASHCSL 399
QY 527 -----LDRAYLVWRM-----YSYSWTLFTCEIEMLHVSSTADVIQHC--ORVXP 571
DB 400 CKSTGIGDVAVATWEGKPTFERQOMSLNNHVFPSKVS--PSLSIQEILQGLKRLSF 457
QY 572 IIDLVHAKVISTDLSADC-----LIPITSRYIMLQRLTVIISP-----VD 613
DB 458 LIDLIGGLVTPNRLATLSSGGEQERAIKHLGSELGFTIYIDEPISGLHPDTEKLIG 517
QY 614 VIASCNCLTVLAARPAVWTDLRHTGFLPFVAHVSSLSQMSISGMAAGYG----- 668
DB 518 VI-----KKLRDQGNVTLVHEBERMISLADRIIDIIGGAGIGFGEGLP 561
QY 669 -----NLANSRPOGEY-----GVTIAFLRL 690
DB 562 NKGPEDFLWSSSLTKYLRQELTIPRPSREAPTSMLLTETATINLKNLSRLPLALU 621
QY 691 ITTLVKGQUGSTOS-----QGLVPCVWFVLEKMLPS--YHKWRYNSHGVREIGCLILELI 744

```

```

Db      622 IG--VTGWSGSGKSLINNTIVPAIESFLKQENPNKMHFEW-----GC-IGLI 667
Qy      745 HAILNLCHETDLHSHTPSLOFLCISLAYTEAGOTVINIGIDVTIDMMAAOPRS-- 802
Db      668 HI-----TRDLPGRSORSI-----PLTYIKA-----FDDRRELFASCPRLSR 704
Qy      803 -----DGA--EGOGGOLIKTKVLAFTVNNVIRLKPENNVSPLE-----Q 843
Db      705 OGLTKAHFENPOGACIQCGGLTMTI-----SDDPTLPICSEQ 745
Qy      844 ALSOHGA-----HGNMLIAVLAKYIY-----HKGDPLPRLAIOLKRLATV 885
Db      746 GKRAHSEVLEILYEGKNIADLDMTAYAEKFPISHPKIKHAKLGLRD----- 797
Qy      886 APMSVYACLDAAAIRDAFLTRLOSKIE---DMRIKVMLEFLTVAETQ--PGLTEL 939
Db      798 -----YLPGLRPLSTLSCGEIQRLLTAHELFPASPKOTLVYLDDEPTGLTHDIOALIEV 852
Qy      940 FLNLEVDGSGSKSPSLGMSCLHAVLELIDSOQODRWCPRLHRAIA-FLHALMOD 998
Db      853 LLSLTY-----LG-----HTVLVIEN-----MHVVKCDVYVLELGPB 885
Qy      999 RRD--SAMLVLRTRKPFMENTLSPFGTLPSPSETS-----EPSILETCALIMKIC 1048
Db      886 GGDLAGYTLASTCRDILQNTPTAKALAPIESLDPVVKSEPPSSPKSCDILIK--- 942
Qy      1049 LEIYVYVKGSLDOSI-KOTL-----KKPSIEKRPAYWSG-----YVKSIAV 1088
Db      943 -DAYONMKHIDLALPRNSLIAGPAGSGKSLVFDILVAGNAYAELEPPYRQGLL 1001
Qy      1089 -----HYAETEGSS-----CTSL-LEYOMLVASW-----RMLIITATADI 1124
Db      1002 KETPLPSYGEVKGSLPVISVRKSSSNRSYHTIASLALGSLNGLEBELFALIDEPSPLEB 1061
Qy      1125 ---MHLTDSVVRROLFDVLDGTAKALLVPAVACILGSMKCTLLILRLQMKRE--- 1177
Db      1062 KLSKIHLPSSSIATQSYNDEL---TSLTSP-----IFLGAD---LEIFQEKQKEGFI 1109
Qy      1178 -----LGSVDEILGPLETEILEGVLDQDQOMEXTKAKVSAFITVLOMKEMVSDIPQ 1230
Db      1110 KLYSEGNYDLDERL-PLNLEIPALVIOHTKVPKSSSLISATSVAPSLSEIWIYISQ 1168
Qy      1231 YSOJLVANVCETLOBEVIALPQTRHSLALGSATEKDMETDDCSRHROGVCYL- 1289
Db      1169 KKQRKLSYSLGCKDKGRLYPEITHQL-LSS-----DHPERGCLTJC 1208
Qy      1290 ---GLHLAKELCEVED-----GDSWLQVTRRL-----PILPTLLPTL 1324
Db      1209 GGRGEILKISLEHKEKIAHTTPEFSLFPFKSYMKVQVGLKDENASQPL-KLTLT- 1265
Qy      1325 EVSLRMKNLHFTTEATLH-----LLTLTARTQO-----GATVAVAGAGITQSI 1366
Db      1266 ---KEFLNFCGSSSEFPKNNALIMBQDLTESDPLIKPLALATCSPACKSGSLANDYA 1319
Qy      1367 -----CLPILSVYQ-----LSTNGTACTPS-----ASRKL- 1392
Db      1320 NYVRINNTSLDIDYQEDATPLESFINTGTDTRSIIDLMLNRLTFISKVGLSYITLQOR 1379
Qy      1393 -DAPSPGCVYRLSNS-----LMEOULLKTRVNFPEALDFVGVQERTLQCLN 1439
Db      1380 QDTLSDGNYHLHLAKKISINLTNIVYLFEBEFLSLHQQDLPITYQLL---KELVANNNT 1446
Qy      1440 AVRTVQSLACEADHTV-----GTLIQLSNEN---KEWHFHLPOLM----- 1478
Db      1437 VIATDRSGSLIPHADHAIPLGSGSGPGLMDSPTEVCPSVDLHANVPQREVCAPLIS 1496
Qy      1479 -----RDIOVNLGYLQACTSLHSRKNLQHTLQNKNGSLPSAAYAKQRPSPS 1527
Db      1497 ISKANHTGSDRTLKVNLSI-----HHIQN-----LKSASALH 1529
Qy      1528 A--ASAPSSSKOPADTEASEQALHTVQYG-----LLKILSKTLAALBHFTPDVCOI 1579

```

```

Db      1530 ALVAGVSGSGKTSLLBEGFKQAEILLAKGTTFSDLVVIDSHPIAS----- 1578
Qy      1580 LLDPSLDLAENRFLPALSTFTPTPDSVAPSGFLTATVN--VALNMLGEL---DKKER 1634
Db      1579 --SQRSIDISTY-----FD--IAPSLAFVASTQAKNLNISTMTFSTNTXQOQ 1622
Qy      1635 LTOAVGSTOAEGRTLKSLMLFTMENCYLL-----ISQAMRY----- 1673
Db      1623 CSDCGGLGYQ-----WIDRAFALERPPPTGSGFRIOPLADEVLYEGKHPG 1669
Qy      1674 -----LRDPAVPRDKORMKQELSELSTL-----LSSLS-----RYFR 1708
Db      1670 ELHLPTEIVALARPPFIKTIQKP-LKALLDIGLYPLIGGLSELSSVSEKALTAYFLY 1728
Qy      1709 GAPSPP 1714
Db      1729 QTPETP 1734

RESULT 31
D96746
hypothetical protein T9N14.23 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: D96746
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hutzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maltli, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96746
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2777 <STO>
A:Cross-references: UNIPROT:Q9C7S8; GB:AB005173; NID:g10645381; PIDN:AA621500.1; GSPDB:GR
C:Genetics:
A:Gene: T9N14.23
A:Map position: 1

Query Match 1.6%; Score 140; DB 2; Length 2777;
Best Local Similarity 18.7%; Pred. No. 2.4;
Matches 359; Conservative 238; Mismatches 609; Indels 716; Gaps 95;

Qy      96 LLOCYLOEDYRGTDSV-----KTYLODERQOALIL-----KIADYVEERCIL 141
Db      1110 LVQSVLSEVDGSKDSCMLPLRVLLFTQSLSEPLHSGRTTCGLADTSFATLDEI 1169
Qy      142 RCVLHLTYFQDERHPVREYADCVDKLEKELVSKYQQ--FEBLYKTEAPWETHGMLMT 200
Db      1170 KGLMRSTI-----PDEVGIVNAFSSALICATESEILKNPASYMWDYXGSF- 1219
Qy      201 EROYSRMFVQCLREQSMLEIIFYAVF---EMASDL---LVTKMKEGQFSGSQ 252
Db      1220 -----SFLSLIIFLEKNPLGNLSKLSL-PFMRLGLELT----- 1251
Qy      253 TNRHLVDETMDPFRDGRGYSPALIVEG-MDIESLHKCALDDREHLQF--AODGLICQD 309
Db      1352 VSRNLREGTVDSRIDPADHSSTTEKISKODIPDIESLAFSVFLEQTFPVLLNLEIWSMD 1311
Qy      310 MDCL---MLT-----FGDI-----PHHAPVL-----L 328
Db      1312 ISCLPEPRRLTELLKLVQSPKSDISIDRLILFMLFOIRSLKYQPH--PYLCQOSEI 1369
Qy      329 AMALLRHTLNPETSTSVVRKIGGTAIQLVNFQYLTR--LIQSLASGNDCTTSTACVCY 386
Db      1370 CLRMRHLFSGISKDLVSGPSADKLKHQVQGVTLSPHVWMALESBADCTLPRVQNV 1429

```

```

QY 387 GLLSFLTS-----LEHITLGNQODIITPACE-----VLADPSLPFLFW 425
DB 1430 VFSELLLNAGRLGISEIDQHIL-----DLASTCEFLPEESHIKRGDLRADKSI----- 1480
QY 426 GTEPISGIGIIDSVCNMF-----PHLISPL-----QLIRALVSKSTAKKVV-----S 470
DB 1481 -----MARKVVERLLVFRDKEFLCVGSOSYAPILQHPLQALIKFISPFKLLYLHNS 1535
QY 471 FLDKMSFYNELYKHKPHDIVISHEDETLMWRQTPKLLYPL-----GGOTNRIQGVIGVQWML 527
DB 1536 MLSK-TYEEEL-----ASPLISFGLDLAGAFEMLLIYSR--QPA 1573
QY 528 DDAVALVWEYSYSMTLFTCEIEMLHVSTADVIQHCCQVKKPIIDLHVHVISITDLSIA 587
DB 1574 KRRVDFMELEEDKNY-----DSRIIEGVSLA-----CR-----FSTRELA 1611
QY 588 D---CLPFTSRIV-----MLORLITTVISPPV-----DVASCVNCLTVLAARPAK 632
DB 1612 SADTCLLVKVGSGIFPGNSOHCVSQHQLTVIMSOIVGRTSKDLIHCIN-----QASMSRAK 1667
QY 633 VMTDLRHGTFLPFVAHPVSSLISOMISAEQMANAGVGNLHNS--EPOGEVGTIAPLRLI 691
DB 1668 I-----LFYIVESSPLHLSVFGHIFFSMLSKLQGDNA-----LI 1701
QY 692 TTLVKGQUGSTQSGVPCVNFVKEMLPZYHKRKNYSHGVREOIGCLILELHAIINLC 751
DB 1702 T-----DDQFVMLPVLPLFLASVFAKLEK-----SC-----SKC 1731
QY 752 HE-TDLHSS-----HPSLOFLCIGSLATTEAGQVYINIGIVDIDIMVMAAQRSD 803
DB 1732 LDTLSLSNIIKIGFLQWPKFCSCGCTFEKYE-----LILST-----SE 1771
QY 804 GAEGGQGGQLIKTVKLAFSVTNNVIRLKPPSNVVSPLQALSOHGAHNNLIVAKYI 863
DB 1772 DIETMFASNLGKAVRM-----FQ 1790
QY 864 YHKIDPALRLAIOQLKRLATVAP-----MSYVACLGNDAAIRDAF--LTRLQSKIE 914
DB 1791 YHFLTESPTKTDLLKVFYMFHTSAGKEMLYEIKEDVVKSVQDFNITNRLVAKVE 1850
QY 915 DMRIKVMLEFLVAVETOPGLIELFNLBYKQSD--GSKESFL-----GMSCL 963
DB 1851 LSRLICLPEDSCMHLKQAGCC-----XKSSPEKMSRBSLNPILNAFVNTQC- 1902
QY 964 HAVELIDS-----QOQDRYV--CEPLHRAIAIFALHMODRDSAMLVLRPKF 1013
DB 1903 -----VVERSGYKXGNSEREODKWFCKSLEY-----FILRSILKE 1941
QY 1014 WENITSPFLGTLSPSETSEPSIIEETCALINKIICELIYVYKGLSDQSLKOTLKRESIE 1073
DB 1942 LEGMCEEL-----AHLDSPFLERIMNLILRYFKDS--KITKILREIFSF 1986
QY 1074 KRFAYMSGYKSLAVHVAETEGSSCTSLLEYQMLVSARMMLLIATTH--ADIMHMLTDSVY 1132
DB 1987 SRGKY-SYHQDVLVSHQFTSISLS-----ISSHTEGVIRIPVSSIL 2029
QY 1133 RROFLDVLDTKALLLVASVNCILGSMKCTLLILLRQWKRELGSVDEILGPLETEL 1192
DB 2030 K-----LTIIP-NLMSVRVEN-CS----- 2046
QY 1193 EGVLANOQOLMEKTKAVFSAFITVLOMKEMKVDIPOYSQVLNVCEILOEVIALFDQ 1252
DB 2047 -----LBAPEYLSQIILIKLIGVLLCKCKDSGIFPKDL 2081
QY 1253 TRHSLALGSAITEDKDSMETDSCSRHRDQDGVCLGLHAKELCEVSD--GDSMLQ 1309
DB 2082 HFRLLCGVATPSSIIDQS-----YKLMHDIULYGEHTIAND--SEDIYMGNAALK 2130
QY 1310 VTRRLP---ILPTLLTLEVLNRMKONLHFE---ATLHLLLTARTQGA--TAVAGAG 1361
DB 2131 IREBLPEDGSYSYDIEDLRQSL-FKENILCLDPKRCQAQTVYFPGRTAASDMTYIYDDP 2189
QY 1362 ITGSLICPLISLV-----QLSTNGTAQTPPSASRSKSLDAPSPGV 1400

```

```

DB 2190 ISEK-CSPALERDPPFILLAVSVQLSMTYIEVERPASGLLAVAPRSMASADL-----GI 2244
QY 1401 YRL-----SMSIMEQLLKYRLNPLPE-----ADDFVGHQERTLOCT--N 1439
DB 2245 RKLGETYGLIFVDLEGIPLEHDFGSSSTVNFSSQRLMVRLVFGLESEDAQIYIRNS 2304
QY 1440 AVRTVQSLAEBA-DHTVGFILQSNFMKEMFHLPLQMRDIOVNLGYL--COACTSL 1496
DB 2305 ILETVMGFSSPLADDETKGLIQVVRKSVKHL-----KARHLVENCGLLSWCSFFSML 2360
QY 1497 HSRKMLQHYLQNKNGDG-----LPSAVQRYORPPSASAAAPSSSKQPAATSEAS 1546
DB 2361 TTK-----PTGDEDSFVVVLEVIDALASR-----NTENS 2392
QY 1547 EQQALHTVQGLKLIKISTALAHHTPDVQCILLDOSDLAEVNFALPST----- 1599
DB 2393 QRSALB-----GLMEISS-----RLYT-----LLEDGLVSMQENGTSLISATLKISHKR 2436
QY 1600 ---TPPDESVAPSPGILLATVVALNMLGELDKKKEPLTQAVGLSTOAGRTTSLIM 1656
DB 2437 KKNQPHFTTIBGIFQULFEAAVNC-----DSPQVEA-----SABG--RLDTILM 2478
QY 1657 FIVNENCFYLLISQAMRYLDPVAV--PRDKORKQELSELSTLSS----- 1701
DB 2479 ST-----PVEIICMDVHRLRRLFMGSSSTALSKDKKSGKPECHQ 2520
QY 1702 -----LSRYFRGASSPATGVLPSPQGSTS--LSKASPEQEPILQVQAFVRA 1750
DB 2521 DTKHTTEETVAKFLMLASVILGKLYSEANDDOIYVSETPEPLPTLLEYLKR--RN 2578
QY 1751 MQ 1752
DB 2579 LQ 2580

```

```

RESULT 32
A57099
DNA-activated protein kinase, catalytic subunit - human
N/Alternate names: DNA-PK-cs
C/Species: Homo sapiens (man)
C/Date: 27-Oct-1995 #sequence, revision 27-Oct-1995 #ext_change 09-Jul-2004
C/Accession: A57099, I59408
R/Hartley, K.O.; Geill, D.; Smith, G.C.M.; Zhang, H.; Divecha, N.; Connelly, M.A.; Admon,
Cell 82, 849-856, 1995
A/Title: DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinosit
A/Reference number: A57099; MIMD:95401275; PMID:7671312
A/Accession: A57099
A/Status: nucleic acid sequence not shown; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-4096 <HAR>
R/Suppl. J.D.; Menniger, J.C.; Hartley, K.O.; Ward, D.C.; Jackson, S.P.; Anderson, C.W
Proc. Natl. Acad. Sci. U.S.A. 92, 7515-7519, 1995
A/Title: Gene for the catalytic subunit of the human DNA-activated protein kinase maps to
A/Reference number: I59408; MIMD:95365397; PMID:7638222
A/Accession: I59408
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1789-2203 <RES>
A/Cross-references: GB:I27425; NID:9667975; PIDN:AAA79244.1; PID:9667976
C/Genetics:
A/Genes: GDB:PRKDC
A/Cross-references: GDB:234702; OMIM:202500; OMIM:600899
A/Map position: 8q11-8q11
C/Function:
A/Description: involved in double-stranded DNA break repair and in V(D)J recombination
C/Keywords: DNA binding; DNA recombination; DNA repair; nucleus; phosphotransferase
Query Match 1.5%; Score 139.5; DB 2; Length 4096;
Best Local Similarity 18.5%; Pred. No. 4.9; Mismatches 368; Indels 433; Gaps 56;
Matches 215; Conservative 148;

```

QY	354	QOLNPFQVLTLLQSLASGNDCTSTACMCYUGLLSPVLSLEHTIGNO-----	404
Db	905	IFLDVF--LPRTTELATPASDQOTVAACELLHSMVMELGATOMPBGQAGAPMYOLY	962
QY	405	-----ODIIDTACEV-----LADPSLPEHF--WGT-----EPTSGGIIIDSV-----	440
Db	963	KRTFFVLLRLACDVOQVTRQLYEPVLMQILHFTNNKKFESQDPTVSLLEAILDGIIVDVPD	1022
QY	441	-----GCMFPHLLSPLL-----QLTRALVSGKSTAKKYSF-----LDKM	475
Db	1023	SLTRDFCG---RCIREFLKMSIKQITPQOQEKSPVNTKSLFRKLYSLALHPAFKRLGAS	1079
QY	476	SFYNELYGHKPHD-----VISED-----GLTMR-----RQTP	503
Db	1080	LAFNNIYEFREESESLVQFVEALVIYMESIALHADEKSLGTIOCCDAIDHLCRITE	1139
QY	504	KLYLPLGGQTNLRIPOG---TWQVWMDRAVLVMEKXSYSWTLFTCEIEMHLNVSTA	560
Db	1140	KKHSVLNAAKKRRLRGRFPFPPSLSCLLD---LVKM-----	1171
QY	561	DVIOHCQEVKPIIDLVKVISTDLSIADCLPEITSRIYMLQRLTTVISPDVIVASCVN	620
Db	1172	-LNAHGR--PQTECRHKSIELFYKVP--LLP-----	1192
QY	621	CLTVLAANPAKVMTDLRHGTFLPVVAHPVSSLQMSIABEKN-----AGYGNLMMN	673
Db	1200	-----GNRSP--NLW-----LKDVLEKBEESFLINTFEGGCG-----	1230
QY	674	SEQPOGGERG-VTIAFLRLITTLVKQGLSTQSGVLPCVMFLKXMLPSYHKMRXNSHV	732
Db	1231	--QBSGIIAQPITYLR-----QPSIQATL-C-----W-----	1255
QY	733	REOIGCLILELHAILMNCETDHSHTPSLOFLCISLAYEAGQVINIMIGIVDTI	792
Db	1257	-----LDLILALAE-CVMTFGERVTGALQVL-----GTEAQSILKAVAFLESI	1300
QY	793	DM--VMAAOP-----RSDGAGGCGGOLLITTVKLASVTNNVIRLKP-----	834
Db	1302	AMHDIIAEKEFGTGAAGNRKTSPOGGERVNSKCTVAVRIMEFTTLINTSPBCKMLKX	1361
QY	835	-----SNVVSPLQOALSQGHAGNNL--LAVLAKYIYHNHDPRLPALIQLKRLATVAPM	888
Db	1382	DLCNTHLKRVLQVTCERPASIGFENIGDVQVMNH-----LPDVCNLMALM---KM	1408
QY	889	SVYACIGDNDAAIRDAFLRLOSKIEDMRIKIMILEFLVAVETOPGLELFLINLEVDG	948
Db	1409	SPY-----KOILETHLREKTIASIEBLCA-----VNIYGEDA	1441
QY	949	SDGSEKESFLGMSCLHAVALLELIDSOQODRYWCPRLLHRAALAFHLAMODRRDSAMVLIR	1008
Db	1442	QVDRSRLAAVNASCKO-----LHRAGL--LHNILPQSGTDLHHSVG	1480
QY	1009	TKPKRWENLTSPLEFGLSPSPSTSPSILETCALIMKIIICLEIYYVVGKSIDQSLKOTLK	1066
Db	1481	TE-----LLSLVYVGIAFDGDERO-----CLP-----SLDLSCKQ--LA	1511
QY	1069	KFSIEKRFAYNSGYVKSU-----AVHVAETEGSSCTSLSEYOMLVSAMRMLLIATTH	1121
Db	1512	SGLELAPAF--GGLCERLVSLILNPAVLSTASLGSSQSVIHFSH--GEYFSLPSEFIN	1566
QY	1122	ADIMHMLTDSVVRQFLPDLVDGKALLVPASVNCRLRGSNK-----CTLLILIR	1177
Db	1569	TELLRNLPLAV--LEIMOSSVDNTK---MWSAVLNMQLQSFPERANOKHQGIKLATTTILO	1624
QY	1173	QMKR-----ELGSVDEILGPLTEILEGVLQADQOUMEXTK---AKVSAPFTV-----	1217
Db	1635	HMKKCDSSWMAKOSPLETEKAAVATALLAKILQIDSSVFNTSHGSPFEVPTTYISILADTKL	1684
QY	1218	-LQMKEMKVSIDIPOYSQVLVANC-----TLQEVNIALFDQTHSIALSASEDEK-----	1266
Db	1685	DULHKGQAVTLLPFTTSLTGSLLEELRVRVLBQLIYAHFPMQSERPFPGRPNNTVDDCKM	1744
QY	1267	---DSMETIDD-----GSHSRH-----RDQDGVCVGLHMLAECEV	1300

```

Db      1745  KFLDALEISQSFMLEFMTEVLCREOQHWEELFOSSFRIARRGSCVTVGLLESYEM 1804
QY      1301  DEDGSMLOVTRRLPILPTLTTL 1324
Db      1805  FRKDDPRLSFTQSFVDRSLTLTL 1828

RESULT 33
A48126      translation activator GCN1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein G1318; protein YGL195w
C:Species: Saccharomyces cerevisiae
C:Date: 21-Jan-1994 #sequence revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A48126; S62050; S64212
R:Matron, M.J.; Crouch, D.; Himebusch, A.G.
Mol. Cell. Biol. 13, 3541-3556, 1993
A:Title: GCN1, a translational activator of GCN4 in Saccharomyces cerevisiae, is required
A:Reference number: A48126, MUID:93268304; PMID:84972629
A:Accession: A48126
A:Molecule type: DNA
A:Residues: 1-2672 <MAR>
A:Cross-references: UNIPROT:P33892; EMBL:L12467; NID:G311099; PIDN:AAA34635.1; PID:G3111
A:Note: sequence extracted from NCBI backbone (NCBI:N:132672, NCBI:P:132673)
R:Klima, R.; Cogilivina, M.; Bertani, I.; Zaccaria, P.; Bruschi, C.V.
submitted to the EMBL Data Library, September 1995
A:Reference number: S62045
A:Accession: S62050
A:Molecule type: DNA
A:Residues: 1-2672 <KLI>
A:Cross-references: EMBL:X91837; NID:G1177627; PID:e203621; PID:G1177633
A:Experimental source: strain FY1679
R:Brusch, C.V.; Cogilivina, M.; Bertani, I.; Klima, R.; Zaccaria, P.; Delneri, D.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64183
A:Accession: S64212
A:Molecule type: DNA
A:Residues: 1-2672 <BRU>
A:Cross-references: EMBL:Z72717; NID:G1322820; PID:e243797; PID:G1322821; MIPS:YGL195w
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:GCN1
A:Cross-references: SGD:S0003163; MIPS:YGL195w
A:Map position: 7L
C:Keywords: translation regulation; transmembrane protein
F:69-85/Domain: transmembrane #status predicted <TM1>
F:216-233/Domain: transmembrane #status predicted <TM2>
F:929-945/Domain: transmembrane #status predicted <TM3>
F:993-1009/Domain: transmembrane #status predicted <TM4>
F:1704-1120/Domain: transmembrane #status predicted <TM5>
F:2021-2037/Domain: transmembrane #status predicted <TM6>
F:2272-2288/Domain: transmembrane #status predicted <TM7>

Query Match      1.5%; Score 138.5; DB 2; Length 2672;
Best Local Similarity 18.0%; Pred. No.2.9;
Matches 301; Conservative 227; Mismatches 523; Indels 621; Gaps 77;

QY      326  VILAAALRHRTINPESTSVVRKIGTALQANFYVITRLIQ--SLASGNDCTTSIRAC- 382
      :|||:::
Db      975  ILLAMEITISVAHAEPDPSIR-----ISIVEVLLSLSPSKAKIARDC-FNALCQ 1025

QY      383  -----MCVVGLLS--FVLTSLSLHTLGNQODIID-----TACEVLADPS 419
      :|||:::
Db      1026  SLSVAPNGDDLMILSNLSPNGFVASTI-LETLNDHELELPPMKYSPEVFICFDSDDS 1064

QY      420  --LDELFWGTEPTSGLIILDSVCGMPPHLSPLIQLLRA-----LVSGKSTAKK 467
      :|||:::
Db      1085  NREIADFIWEFKFVVNDLKLSPFLFNQDDSG-LRLFAANVAVGAVSLFTSEENSSK 1143

QY      466  VSPFLDKSGFVNEELVGHKHDVSHEDGTL-----WRQD-----TPKLL---YPL 509
      :|||:::
Db      1144  DV-LNDLNLVYKE--KAKPLLEPLIDPGLVLVASSEKQDPMQGSSTVAITLTKIAKAFSA 1200

```

510 GGGTNRIRPGTGVGVMDDRAVLVWWEYSXSWTLFTCEIMLHVSTADVIQHGRV 569  
DB EDDTVVNIKPLVDDGGLVDBEPIRQEMKRGAVLIT-----LHGSNS 1245  
QY 570 KPIIDLVKVI--STDLSIADCLPI-----TSRIYMLQR-LTTVISPPVDVI 615  
DB 1246 KDLPFEEALSSSTDSALKENVILLYGTARHLQQSDARHTHTLERLSTLTPSADIQ 1305  
QY 646 ASCVNCCLVLAARNPAKVTDLRHTGFL-----PFVHAPS-----SLQWISAEQNA 664  
DB 1306 QAVSACIPLVLFQFKQKVD--YGIIMKELNFTVASSMRKGAANGIAGLVGYSISA 1362  
QY 665 GGYGNLNMN-----SEPOGEYVITAFRLITTL----- 694  
DB 1363 LSEFDIINHLIEAADKKEPKRRESVGFQYLSLSLCKFPPEYIELPMLNKGAV 1422  
QY 695 -----YKQUGSTQSGO--LVPQMFVLKEMLPYHKRPNYSHGR----- 733  
DB 1423 PEVDATATARKATMAHTTGQVKKLIPVAVSNLDEI-----AWRTKGSVQLGNMAYL 1477  
QY 734 -----EQIGCLIELHAIINLCHETDHS-----HTPSLQFICISL----- 772  
DB 1478 DPTQLSASLSTIVEPIV--GVLDNSHKEVKADESLSKRFGEVINPBIQKLVPLLAIG 1536  
QY 773 --AYTEAG-----QT-----VININGIGV 789  
DB 1537 DPTKYTEALDSLQTOFVHYIDPSLALIHIIHRGHNSANIKRKAKCIVGNMILV 1596  
QY 790 DTIMWMAAO-----PRSGAGGQGGOLLITVTKLASVTNNVIRLKPPS 835  
DB 1597 DTKLIPYLOQIDVEIAMVDVPNTATATARALGLAVERLGEQF-----P 1644  
QY 836 NVASPLLEQALSOHGAGNNL--IAVLAKYI--YHKIDPALPLAIQLKRLATVAPMS 889  
DB 1645 DLIRLDDTLSDSKSGDRLSAQALAEVIGLITKIDEMLPIT-----LGVNFR 1697  
QY 890 VY-----ACIGNDA--AIRDAFLRLQSKIEDKRI-----KXMIIEPL 926  
DB 1698 AYIEGFWPLLLFLPVCFSQGFAPYINQIIPILSGLANDENIRDTALKKGLIVQYA 1757  
QY 927 TVAV--ETOPGLIE-----LF-----LNLVKDSDGSKERSLG 958  
DB 1758 TKAVADLLPDERGMFENDRIRLSVOLGELLFQVTGSSNRNFESEDDNHGERS-- 1815  
QY 959 MMSCLHNVLEIDSOQODRWYCPRLHRAAIAFLHALMODRDSAMVLRTKPFEMNLT 1018  
DB 1816 -----GKLVDYLGQDRDR-----LLALFVCNDSIGVATYDWMK--- 1854  
QY 1019 SPLFGTISPESETEPSTLETICALIMKICLETIYVVGSLDGLKDTLKKFSTIEKRFAY 1078  
DB 1855 -----ALVPNTPRAVKEILPTLQGITVTHLASSNVLRNIMAOGLDVLVRVGNA--- 1905  
QY 1079 WSGVYKSLAVHVAETEGSS-----CTSLLEKQMLVSA-----WRMLLIATTHADI 1124  
DB 1906 LSQLPLIEESLITNSDSRGVCIAL--YELIESASTETISOPOSTIVAILITLALID- 1962  
QY 1125 MHLTDSVVR-----QLFLDVLDTKALLVPAVAVCLRGSKYKTLHLLRQMKREL 1178  
DB 1963 ---ESATVREAAALSPFVQDVNGKT----- 1985  
QY 1179 GSVDEILGFLTEIEGVLQAD-----QQMKTAKAVFSAFITVL----- 1218  
DB 1986 -ADEVLPYLIHMLSESDNSDPALLGQELMSKSDVIFPLILPTLLAPPIDAFRASALG 2044  
QY 1219 QMKEMKVSDFIROYSQLVLANVCELTQEEVIALFDOTRHSIALGATEDKDS-----METDD 1273  
DB 2045 SLAEVASALYKRSIIIN-----ALVD-----AIGTSBESKGALELALDR 2088  
QY 1274 CSRRHRDQBDGVCVLGLH--LAKELCEVDEBDGSMQVTRRLPLTLLTLLTLEVSILMAKQ 1332  
DB 2089 VFLSVNDE-----GLHPLLQOIMSLKSDN-----IEKRIAVLERLPHFPKXTV-LDF 2136  
QY 1333 NHTETATLHLLTLARTQOGATVAVAGITQISICLPLSVYQSLSTNGTQTPSARKSL 1392

DB 2137 DVYIPNFVSAIISLDEDDR-----VNGNFMNALSTLLKKV 2173  
QY 1393 DABSWPGVRLSWSLMEQLLKTIRLYNFPALDFGVGHORTLOCLNAVTVOSLACEE 1452  
DB 2174 DKPT-----LEKLVKPAK-----QSLALYGR 2194  
QY 1453 ADHTVGRILQLSNFMKEMHNLPLQMRDIOVNIQVLCQACTSILHSKMKLQHVLOKNGD 1512  
DB 2195 QGQDVA-----AKLPR-----GPNCLV-----PIFLHGLMGSND 2225  
QY 1513 GLPSAVAQVRPPSAASAPSSSKOPADTE-----ASEQALHTV 1554  
DB 2226 -----ERESALATADVSKTPPANKRFVSVITGPRLRVNGERSSDIKA--AI 2273  
QY 1555 QYGLKILSKTTLALRHFPTDVCQIILDDSLDAEYNFLFALS-FTTPFDSEVAPFET 1613  
DB 2274 LFLALNVLFKIPMFLPFLIPOL-----QRTFVKSISDANVETLRLRAAKALGA 2321  
QY 1614 LATVAVNALMGLDKKKEPLT-QAVGUSTQA--EGTRT--LKSILMFTMB 1660  
DB 2322 LI-----EHQPRVDPLVIELVTGAKQATDGVKTAMKALLEVIWK 2362

## RESULT 34

S46715

hypothetical protein YHR099w - yeast (Saccharomyces cerevisiae)

N/A:alternate names: Hypothetical protein H9332.1

C/Species: Saccharomyces cerevisiae

C/Date: 28-Oct-1994 #sequence\_rev150n 28-Oct-1994 #text\_change 09-Jul-2004

C/Accession: S46715

R/Vaudin, M.

submitted to the EMBL Data Library, May 1994

A/Description: The sequence of S. cerevisiae cosmid 9332.

A/Reference number: S46715

A/Accession: S46715

A/Molecule type: DNA

A/Residues: 1-3744 &lt;NA&gt;

A/Cross-references: UNIPROT:P38811; EMBL:U00060; NID:g487928; PIDN:AB68923.1; PID:g4879;

C/Genetic:

A/Gene: SGD:TRAI

A/Cross-references: SGD:S0001141; MIPS:YHR099w

A/Map position: 8R

Query Match 1.5%; Score 138.5; DB 2; Length 3744;

Best Local Similarity 17.6%; Pred No.5;

Matches 252; Conservative 214; Mismatches 492; Indels 477; Gaps 61;

QY 100 YLQEDYGRTRDSYKTVLQD-----EROSQALILKADYYVERTC 139  
DB 317 FLQDYVNFVBDLIRLLQDCPSLSSARKELLHATRHILSTVYKGLFLPKLDYLPDERIL 376  
QY 140 I-----LRVYLHLITTFQDERHPYRVEYADCYDKLEKEIVSKYRQPFELYKTEAFTWET 194  
DB 377 IGGFTMHETRLRAYL-----STVADFHNIRSEL-----QLSEIRKTX 414  
QY 195 HGNLMTERQVSRNFVQCLREOSMLLEIFLYAAVFEAPMSDILLVTKMKQEGFSGRQTN 254  
DB 415 -----IKIYGYLDESLALTVOJMSAKLLNVLRIKLGKKNPQEPAPRAKKLL 464  
QY 255 RHLVDETMDFVDRIGYFSALILVGM-----DIESLHAKALDRRLHOPADGLIC 307  
DB 465 MLIIDSYNRFKILNRQYDTIMKYGRVETHKKERKELKNSIQDNKDESEFMRKVLBP 524  
QY 308 QMDICM-----LTFGDIPIHAPVILAMALLRHTLNPEE 341  
DB 525 SDDHLMFPQPKEDINDSPDVMTESDKVKNQDVEMFDIKYAPILL-----PTP 575  
QY 342 TSSVVRKIGGTALQNLVFOYLTRLQSLAS-----GANDCTTSTACM----- 383  
DB 576 TNDPIK-----DAF--YLKRTLMSPLKTIHDLKVNPNPNNETTVANPFLMASVSRV 625  
QY 384 -----CVYGLLSFV--LTSLELHTLGNODIITDACEVLADSLPELFWG 426



```

Db      626 FSYBEVIVKDLPHRCITIGLKPFKDHNEKSPETTKKHFDI-----SWRSLP----- 672
Qy      427 TEPTSGGIIIDVSCGFPHLLSLPDLRLALVSGSKTAKVYFLLDKM-----SFYNE 480
Db      673 -----VSATKDARELMYLLFMFMQMDNAFENE 700
Qy      481 LYKHK---PHDVISHEDGTLWRRO---TPKLLYFSGGTNRIRIOG---TVGOVWMDRA 531
Db      701 IIEBELPVTYERMEDEGLHVAQSFLTSETSPNFAGILRFKGLKDKLGNVDFTSN 760
Qy      532 YLVA-WEXYSWTLF--TCEIEMLHVASTADYIOHCQVAPRIIDLHVKVISDLSIAD 588
Db      761 VLILFLKLSFMSVNLFPNINEVLLPH-----LNDILNLSKXSTTAAE 804
Qy      589 CLLPITSRIYML-----QRLTVISPPVDVYASCNCLTVLAARP--AKV 633
Db      805 PL-----VPEYLLRTLFRSIGGRFENLYSIRPIIQLVLLQSLNQW--ILTARLPHEREL 857
Qy      634 WTDLRHT-----GFLPVAHP--VSSLQW--ISAEQNA----- 664
Db      858 YVELCTIVPRLSVLAPLPLMKPLVLPALQQYDPDVSQGLRTLELCTIDNLTAIFYDPII 917
Qy      665 ---GCVGNLMSNEBOPQ--EYGVIAFLRLITLVKQGLSTOSQGLVPCWMEVLKEML 719
Db      918 EPVIDVSKALFNLLQRPFNHAIASHNVRL-----QKLGRRRQFLKRPDTDLTEKTEL 972
Qy      720 PSYHKRWYSHGVBEQIGCLILELIHALNL--CHETDLSHSHPSLOFLCISLA--- 773
Db      973 DIDADPFKINGMEDVPLSVTPGQSALNIIQSYKDIHYKSAVYKVLTCVLLMTKXS 1032
Qy      774 ---TEAGQYINIMGIGVDTIDWMAAPSDGEGGQGLIKYKLAFSVY-- 825
Db      1033 AEPFTNTTELKTAVNSIKERIGIEKNFDELPYKNKDYSGNQLFLRLLESVFYATSI 1092
Qy      826 ---NNV-----IRLKP-----SNVVSPLERQA 844
Db      1093 KELDDAMDNLNLDHFCLLQWTTLLKNKNGNTFNIDIKNRPMLDLSIIDLAPFA 1152
Qy      845 LSGHAGNNLLIAVLAKYIYHK-----HDPALPLAIOQL-----KR-- 881
Db      1153 LSYIIPREVGVALYKRIYEKSCLIYGEELASHFIPELAKOFIHLCYDETYVNRKG 1212
Qy      882 ---LATAPMSVVA-----CLGNDAAIRDAFLRLQSKIEDMKIMLEFLVA 929
Db      1213 VLGIRKLDNVKSSSVFLKTYQYNLANGLFVLKDTQSEAPSAITDSAEKLLI--DLISIT 1271
Qy      930 VETQGLIEFLINLEVKGDSGSKFSLGMSCLHVALELIDSOQODRYWCPRLLHRAI 989
Db      1272 F-----ADVKEEDIGNKYLEN---TLTDIVGELSNANPKVYNAQKSLH--TI 1314
Qy      990 AFLHALMODRDSAMLVLRTRPKFMENTSPFLGTLSPSETSEPSILETCALIMKI-- 1046
Db      1315 SNL-----TGIPIYKLMDSKQFLSPIF-----AKP--LRALPFTMOIGNV 1354
Qy      1047 ---ICL-----EYVVVKGSL-----DQSLKOTLKKS-----IEKFA 1077
Db      1355 DAITFCISLPNTFLTFNEBELFRLLQESIVLADABESISTNIQKTEXSTSEQLVQLIA 1414
Qy      1078 YMSGVKSIAVVAETE--GSSCTSLLEYOMLSVARMMLIATTHADIMHLLTDSVVRQL 1136
Db      1415 ---CIKLLAIAKNEEPATAOQGNIRIRILAVFKTML--KTSPEIINTTYALKSKSL 1467
Qy      1137 PLD-----VLDTKALLLVPAVNCRLGSMKC--TLILLRLQMKRELSVDEILGP 1187
Db      1468 AENSLKPEKLEONGKAPLIMNLSDHQKLTVPGLDLSKLELILAYFKVEIGR--KLDDH 1525
Qy      1188 LT-----EILGCVLADQQLMEKTKAPPSAFITVLQMKEMKVSIIPOYSQVLVN 1237
Db      1526 LTAACRVAVLDTLF--GODLAEQMPTKIIVSINIIFH-----LPPQADMFVN 1571

```

RESULT 35

S09811  
 hypotheical protein U148 - human cytomegalovirus (strain AD169)  
 A:Alternate names: hypotheical protein HFRFO  
 C:Species: human cytomegalovirus, human herpesvirus 5  
 A:Note: host Homo sapiens (man)  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 09-Jul-2004  
 C:Accession: S09811  
 M.:Chae, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerny, R.; Horenell, T.; M., Burrell, B.G.  
 Curr. Top. Microbiol. Immunol. 154, 125-169, 1990  
 A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus  
 A:Reference number: S09749; MUID:90269039; PMID:2161319  
 A:Accession: S09811  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-2241 (CHE-)  
 A:Cross-references: UNIPROT:P16785; EMBL:X17403; NID:G59591; PIDN:CA435407.1; PID:gl7808;  
 A:Note: this sequence was submitted to the EMBL Data Library, December 1989

Query March 1.5%; Score 138; DB 2; Length 2241;  
 Best Local Similarity 19.6%; Pred. No. 2.3;  
 Matches 295; Conservative 190; Mismatches 529; Indels 494; Gaps 71;

```

Qy      398 LHTLGNQODIIDTACEVLADESLPEL-----FW-----GTE-----PTSGLGIIL 437
Db      341 LPALGRYQQLVDEVEGELKALTLPPLPANTSAMTLHAAGTESGANAATATAPSPDEAFLT 400
Qy      438 DSVCGMFPHLL-----SPLLQRLALVSGSKTAKVYFLLDKMSFYNELYKH 484
Db      401 DRLQQLIIHVWQKSCLRPCGPOSAQAQAVRAVL---GLSKLDAFLNWN-LHHGLDLO 456
Qy      485 KPHDVISHEDGTLWRROTPKLLYFSGGTNRIRIOGTVGOVWMDRAVLYWEXYSYSW- 543
Db      457 RMHDYLSHKT-----TKGYSTIDRALLEMQVVPDPYG 490
Qy      544 ---TLTCEIEMLHVVS--TADVIOHCQVAPRIIDLHVKVISDLSIAD---CLLP 593
Db      491 RQHGPALIAWEENLRVYESKPTWELSGRLQRF-----YKRPMPVSDVFCRLP 541
Qy      594 TSRITMLORLITVYISP-----VDVYASCNCLTVLA 626
Db      542 D-----FORLTQVIEGRVRLQROREBYHGVENHLAGLITSIDIHDLDAEDLNRRELTK 595
Qy      627 ARNPA-----KYMTDL-----RH--TGFLPFAHPVPSISQ 655
Db      596 ALQPLDNDNAQOELRLGNAKKLEIOMDLDRISTQULRLVANHIIANGFI-----PVEDLKQ 650
Qy      656 MISAEQWAGYGNILMNSBQPOGEYGVIAFLRLITLVKQGL--GSTOSQGLVPCWMEV 714
Db      651 M-----ENVVEQVLLFPYDLNRKLCDSGYEGGFV-----V 681
Qy      715 LKEMLPSTYHKRWYSHGVBEQIGCL--ILELIHALNLCHETDLSHSHPSLOFLCISL 772
Db      682 IREQD--SY-----LMTGVTRDVNVPRLLOEITLQRLHAYQOATQONERLTOIHDLHVI 732
Qy      773 AYTBAGQTVINIMGIGVDTIMVNA---APRSDGABGQGGQLIKTVKLAPSVTNV 828
Db      733 ---ETLVADPSSGSAITLAVQEQALDELAGGLQLEPVEQORIONAQLAS----- 781
Qy      829 IRLKPPSNVVS-----PLEQALSGHAGNNLLIAVLAKYIYHKHDPALPLAIO 877
Db      782 -RLVBEHEETORFLDGLSYDDPRNEQTIKH-----POLREM 817
Qy      878 LLK---RLATV--APMSVYACLGNDAA--AIRDAFLRLQSKIEDMKIKYMIIEFLVA 929
Db      818 LRRDEQFTLRILINAVLSWFTLVNRLARDESPRTFFDAVSLIAQOLPPDSHEREDLPA 877
Qy      930 VETQGLIEFLINLEVKGDSGSKFSLGMSCLHVALELIDSOQODRYWCPRLLHRA- 988
Db      878 NATYAQWVKLEQIEKAGTASERF-----QALRELVYFRNNEYFPQHWGRIGV 929
Qy      989 ---IAFLHALMODRDSAMLVLRTRPKFMENTSPFLGTLSPSETSEPSILETCALIMK 1045

```



Db 930 GPOVTELYERYOHENEBOH--LERLEREMOEBAGLIVT-----SVEDVQVRVLAR 977  
 Qy 1046 IICLEIYVYVVGSGSDQSLKDTL---KKPSIEKRFAYVSGYKSL---AVHYAETGSSGCT 1099  
 Db 978 APSHRVHMQOQDTLTKKQDFLDKREKROEQROLDGYQKQVQODLQRVDAVKGEML 1037  
 Qy 1100 SLEBYOMLVSAAMRL-----IATTHADIMHLDTSVVRQLFLDVLDTGALLVLP 1151  
 Db 1038 STIPHQPLEATLLELLGLDQRAQPLDFNQDLISALQQLSK-----LDG----- 1083  
 Qy 1152 ASVN-CLR---LGSN--KCTLLILLRQWKELSSVDEILQPLLEILE-----GLVQA 1198  
 Db 1084 -RINECHLGVLTGVVERCHHREAAQTOASLNLHLDQILGPOLLIHETQALQHAHYQA 1142  
 Qy 1199 DQCMETKAKVFAFIVLQMKEMKVSIDIPOYSQVLNVGCELTQOEVIATFDQTRSLA 1258  
 Db 1143 --QTEKQOQDPTTATIGSEFE---GDPARYSSQCKMEQLOE-----TROQMT 1188  
 Qy 1259 LGSATEDKDSMETDSCSRHRDQDGVCLGLHLAKELCEVDEBDGSMLOVTRRLPLP 1318  
 Db 1189 ETSRLDR-SLRQDPGSSSVTRVPEKP--FKGQELAGRITPPRADFQ-----PVFK 1237  
 Qy 1319 TL-----LITLESILMKONLHTEATLHLLTLARTQGAIRAVAGAG 1361  
 Db 1238 TLLDQADARKALSDADLLNQKQVQTLRQ-----RDEQLSTA----- 1276  
 Qy 1362 ITGSIPLLSVYOLSTNGTAQTPSASRKSIDAPSPGVRLSMSLEQLKTLRYNPLP 1421  
 Db 1277 --QVLMTDLVTRHMGSGGLDVTTPDAK-----ALMEKPLETIR----- 1312  
 Qy 1422 EALDPGVHQRRTLOCLNAVTVQ-SLACLEADHTVGFIQLS-----NFMKEMH-H 1473  
 Db 1313 ---ELGKATQO--LPYLSAERTVWMLAFLEEA-----LAQITADPTPHHGSRITYRN 1362  
 Qy 1474 LPOLMRDIOVNLGYLQO---ACTSLHSRKLQYHLQNKNGDGLP-----SAVQORV 1522  
 Db 1363 LQOQAVEAVATLQIENQACENFT---AQHOEATANGASTPRVDMQAVEAWQRL 1417  
 Qy 1523 QRPESAASAPSSSKOPADTEASEQOALHTVQYGLKILSKTALAALHFTPDVCQILLD 1582  
 Db 1418 E-PRVAVGA-----ARHOKVQELLQRLQGTIGDLE-----LQ 1449  
 Qy 1583 QSLDLAEY-----NFLPLASTTP-----TPDSEVAPSPFGILLATVNV----- 1620  
 Db 1450 ETL-ATEYFALLHGIQTFYSYGLDFRSQLEKIRDLTRFAELAKRRGRLSNEGVLPNPRK 1508  
 Qy 1621 --ALNMLGELDKKEPRLQAVGLSTQAEGRFTLKSILMFTMENCFYLLISQAMRYLADPA 1678  
 Db 1509 PQATTSIGAFTRGLNALERHVQLGHQYILNKLNGSSLVYRLLEDIPSVL---PATHTDPA 1565  
 Qy 1679 VHPRDQOR 1686  
 Db 1566 LHMDRRLR 1573  
 RESULT 36  
 T13949  
 neurofibrin - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
 C:Accession: T13949  
 R:The, I.; Hamnigan, G.E.; Cowley, G.S.; Reginald, S.; Zhong, Y.; Gussella, J.F.; Harihar  
 Science 276, 791-794, 1997  
 A:Title: Rescue of a Drosophila NFI mutant phenotype by protein kinase A.  
 A:Reference number: Z17826; MUID:97277221, PMID:9115203  
 A:Accession: T13949  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-2764 <THE>  
 A:Cross-references: UNIPROT:O01399; EMBL:L26502; NID:g1929432; PID:g1929433; PIDN:AA568  
 C:Genetics:  
 A:Gene: NFI  
 A:Cross-references: FlyBase:FBgn0015269

Query Match 1.5%; Score 137; DB 2; Length 2764;  
 Best Local Similarity 18.8%; Pred. No 3.9; Mismatches 627; Indels 602; Gaps 93;  
 Matches 340; Conservative 243;

Qy 169 LEKELVSKYQOEELKYTEAPTW-ETHGNLMTERQVSRWFVQCLRQSMLEIIFLYYA 227  
 Db 755 LQKRIPTLIL-KLEHCVHGVPAWEEFTFRWVESSKYLQYRCKGSDG----- 803  
 Qy 228 YFEMAPBDLLVTKMFEQGFGR---QTRNLHDETMRPVDRIQYFALILYEGMDI 283  
 Db 804 -----AEVF-HRWGKRASHOSSEHDLQEIENWAMTWFLALGV----- 845  
 Qy 284 ESLHICALDRREIHQPAQGLICQMDLMTFFGDI PHAPVLLAWALLRHTLNPEETS 343  
 Db 846 -CHKKS-SRQWMLQOSQNN-----ASLG-----SLQNSLYSSTS 881  
 Qy 344 SVYKIGGTALQLN-----VFQYLTLLQSLASG-----G 373  
 Db 882 SGHSLHPSIVSLSTLPPAPQDVSYCPYQVQGLRLILVCSNEKILNIQKVKEVLG 941  
 Qy 374 NDCITSTACMCYGL-----SVLVSLEHTLGNODIIDTA 411  
 Db 942 EEMSTQ---LVPILPDQVRAIVEKFFDQQGVNNAVNTDINTQFIHTIYIMKSIDPK 996  
 Qy 412 CEVLADPSLBEFWGTEPTSGILIDSVCGMF-----PHL-----LSPILQL 455  
 Db 997 AN--KDPNNQ-----PSRSEHLG--TSIEGMVLGIVRYVRHIDMTYAIRIKTKQLV 1048  
 Qy 456 RALVSGSKTKKVVYSLDKKSPYNELYKHKPHDIVISHEDGTLNRQOTPKLLYPLGGO-- 512  
 Db 1049 EVMKRRDD---LAFQEKFRNKLVEYL-----TDWVGWGSQIAPSSSDAA 1094  
 Qy 513 --TNLRIPQGVQWMLDRAVLV-----EMEYSSWTLFTSGIEMLHNVST 559  
 Db 1095 ILNTSLIFRDLQACHEAVALLRGLPLOPESDRKDLMDASALFLKFTFLPMNLN- 1153  
 Qy 560 ADVIYHQRVYPIIDLVHKVISTDLSIADCLPITSRYMLQRLTVVISPPVDVIASCV 619  
 Db 1154 -DCIDSEBAKEMNNP-----LPPRRP--MAAKGLTLARN----- 1187  
 Qy 620 NCITVLARN--PAKYWTDLRHT--GFLPFV---AHVYSSLSQMS-----AEGNN 663  
 Db 1188 --ATILAMSNLIGANIDSGIMSIDIGYNPDLOTRAAFMEVLQIIOQGFEPFLMETVL 1245  
 Qy 664 AGGVGNL--LMNSEQOGGEVGTIAFLRITTLTVKQGLSTQSGVLPCCMFLKEML-P 720  
 Db 1246 ADRFEQVLQVLTWISDKGELPIMALLANVVTISQMDLAR-----VLTLLPDAKHILSP 1299  
 Qy 721 SYHKWRYNSHGVREQIGCLILELIIHALMLCHETDLHSHHTPSLOFLCISLAYTEAGQT 780  
 Db 1300 LMMNPFREVEVSD-----CMQT-LFRGSLSGKIMAFPKIY----- 1336  
 Qy 781 VINIMGIGVDTIMWMAQPRSDAGQGGQGLITVTKLAESVTNNVITLKPSSNVSP 840  
 Db 1337 -----GASYLQMLLEPLIR-----PLIDBEETCFEV--DPAILEDIEDI--- 1374  
 Qy 841 LEOALSQHGANGNLLAVLAKY---LYHKHDPALPRLAIG-----LYRLATVAPMSV 890  
 Db 1375 -EQ-----HRNNLLATQKVVDALINSSDLRPLQRLSKCHCLYVLSGRFPYL----- 1421  
 Qy 891 YACIGNDAAIRDAFLTRL-----OSKEDMEIKWMLLEFLTVAVE 931  
 Db 1422 --LQNNIGAVGVIFLRFINPAIVSPQELGIVDKOVHSAKKGMLMSKILQNIANHVE 1478  
 Qy 932 TQGGLELFLNLEVKGSDGSKFSLGMS--CHAVLELIDSOQORVCPRLHRAAIA 990  
 Db 1479 PSKEQHMLCFNDLRDHFEGRRFFIQLASDC-----ETVDQTSN--SPISDANVL 1529  
 Qy 991 FLH-AW-----QBRDSAMVLVLRKPKFMENTLSPFLGTLSPSPSETSEPSILET 1039  
 Db 1530 ALHRLMLTHQEKIGDYLSSRDHKAVGRRPDKM-----ATLLAYLGPRPHKVDSHMMF 1584

QY 1040 CALIMKICIEIYVVGKSLDQSLKD-----TLKPSI-----EKFA 1077  
 Db 1585 SS-----YARMSIDMSSTNFEEIWKQMKHEKPKTKLSNNITVQAGTSSG 1633  
 QY 1078 YWGVKSLAVHVAETEGSSCTSLIEYQMLV-----SARMLLIATTHADIMHILTDS 1130  
 Db 1634 YPVVYIARRYKIGETNG-----DLIIYHVIITLKPCFCSPEEVVIDPHTCSIDNRFRTF 1689  
 QY 1131 VVRQQLFDVL--DGTAKLLLVPAVSVNCLRGSMKCTLLILLQMKELGSV--DELIGP 1187  
 Db 1690 LQKMPYVLPVAYENVAHVAI--YNC-----NSWREYTKPHRILAP 1730  
 QY 1188 L-----TEILEGVLOADOQ-----LMEKTKAVFSAPF----- 1215  
 Db 1731 LKGRKLLFLESFPMKLTDFIDAEQKLPGLTSLDEDLKVSNNALKLSKDTKALIKGP 1790  
 QY 1216 TVLQMKEMKAVSDIPQYQVLNVCELTQOEVIAPFDQRHSLALGASATEDKDSMETDGS 1275  
 Db 1791 TALQITSAEKTKVLASHVTLNDVYVASEIEEVCLVDNQFTLSITNES--GQLSFIHND- 1848  
 QY 1276 RSRHRDQDGVCGVGLHLAK--ELCEVEDGDSMLQVTRRLPIIP-----TLTLTBVSL 1328  
 Db 1849 -----DNIVQAIHTRNMEWSQPD-----SVTVHOKIRPKDVPGLLNNALINT 1893  
 QY 1329 -RMKONLHFEATLHLTLTARTOQATAVAGAGI--TQICLPILSYVQLSTNGTAQTPS 1386  
 Db 1894 GSCCPNLR--TNAVNLGALTRTP--LKISQLLFETGGLCP-----SNNTIIFKS 1941  
 QY 1387 ASRSLDAPSWPGVYRSLMSIMEOLLKTLRYNPLPEALDFVGHQERTLQCLNAVTVQS 1446  
 Db 1942 VSEKL--ATNEP--HLTFLEESIQGF-----QRTIEIKH----- 1974  
 QY 1447 LACIEADHVTGFIQLSNEFKEMFHLPOLMRDIQVNLGYLCOACTSLHSRKLQHYL 1506  
 Db 1975 -LCLE-----YMTWP-----LKNLVKCKSNDKSKLKYSOI 2005  
 QY 1507 QNKGKDGIPSAVORVQPPAASAPSSKOPADTASEQALH--TVQYGL-----LKI 1561  
 Db 2006 LDK-----LINITQKEMYPVQAKIMWSIGQIPBLIMVLDNPLHKSITTYGLGSPQEI 2061  
 QY 1562 LSKTLAAL-----RHFTPDVCOIL-----LDQSL--DLAEY--NFLFALSF 1598  
 Db 2062 MADTAVVALASNVQVLSKVVITRICRVMDKSCCTNPQYCLEQHMMDIDAILGRYLLMSF 2121  
 QY 1599 TTPFPDSVAPSFGLATV-----NVALMMLGSLDKKBPLOAVLS 1642  
 Db 2122 NNCL--DVATSVPLTFHTITFLVCSGSPSRKASTHGLVINIHSICTNP-----S 2171  
 QY 1643 TQAGSTRILK--SLMFTMENCEFYLLI-----SOMARYLDPVAVHPRDK-----ORMKQEL 1691  
 Db 2172 FSEBQVRLSLDEBFSLPK--FYLLFGISKVKSAAVTAFRSSCRHPTDKMLGNEKVTOP 2230  
 QY 1692 SSELSTL--LSGL 1702  
 Db 2231 PADRERLSLPSL 2242  
 RESULT 37  
 T13947  
 neurofibromin - fruit fly (Drosophila melanogaster)  
 C/Species: Drosophila melanogaster  
 C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
 C/Accession: T13947  
 R/Title, I.: Hannigan, G.E.; Cowley, G.S.; Reginald, S.; Zhong, Y.; Gueella, J.F.; Harihar  
 Science 276, 791-794, 1997  
 A/Title: Rescue of a Drosophila Nf1 mutant phenotype by protein kinase A.  
 A/Reference number: Z17826; MUID:97277221; PMID:91515203  
 A/Accession: T13947  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Reidses: 1-2802 <THE>  
 A/Cross-references: UNIPROT:O01398; EMBL:L26501; NID:g1929430; PID:g1929431; PIDN:AA589  
 C/Genetics:

A:Gene: NF1  
 A:Cross-references: FlyBase:FBgn0015269  
 Query Match 1.5%; Score 137; DB 2; Length 2802;  
 Best Local Similarity 18.8%; Pred. No. 4;  
 Matches 340; Conservative 243; Mismatches 627; Indels 602; Gaps 93;  
 QY 169 LEKELSVKROQFELYKTEAPTV--ETHGNMTERQVSRWFVQCLREOSMLLEIFLYYA 227  
 Db 755 LQKIMTLIR-KICHCHGVQPAHEETFRWVSSKYLQYPKKGEDQ----- 803  
 QY 228 YFENAPSDLLVLTWPKFQGFGR--QTRHLVDETMDFVDRIGYFALLIVEGMDI 283  
 Db 804 -----AAVF--HRGMGKRASHQSEHDLQEDINEMAWMTFLLAGV----- 845  
 QY 284 ESLKCALDRRELHQRQDGLICQDMCLMTGDDIPHAAPVLLAALLHNTLPETS 343  
 Db 846 -CLHRS--SRQMLQOSQNN-----ASLG-----SLQNSLYSSSTS 881  
 QY 344 SVARKIGGTALQLN-----VFOYTRLLQSLASG-----G 373  
 Db 882 SGHSLHPSTVSLTLPAPRPQDVSYCPVTOFVQLRLVCSNEKIGLNQKVKELVG 941  
 QY 374 NDCSTTACMCVGL-----SFVLTSLLEHTLGNQODIIDTA 411  
 Db 942 EEMSTQ-----LYPILFDQVRAIVEKFFDQGVNNAVNTDINTQFIEHTIYIMKSLDPK 996  
 QY 412 CEVLADSLPELFMGTPTSGIGIILDSVCMF-----PHL-----LSPILQL 455  
 Db 997 AN--KDPNDO-----PSPSEHGV--TSIEGMLGIYRVYHLLDMTYAIRIKTKLCOLV 1048  
 QY 456 RALVSGSTAKKVSFLDKMSFYNELYKHKHDVISHEDGTLMRPQRPKLYPLGGQ--- 512  
 Db 1049 EVMKKRKD--LAFROEMFRKLVYI-----TDVWNGISHQALPSSADAA 1094  
 QY 513 --TNLRIPOGVGVMLDDRVLV-----RMEYSVSWTLFTCEIEMELHAVST 559  
 Db 1095 ILTNTSLFRDLQACAEVALLRGLPLQPEESDRGLMDAKSALFLKYFTLFLNLLN- 1153  
 QY 560 ADVIQCQRVPIIIDYHKVISTDLSTADCLPITSRTKRYMLQLTYIVISPPVIVASCV 619  
 Db 1154 -DCIDSSAEKEMNTP-----LLPRPR--MAAGKLTALRN----- 1187  
 QY 620 NCLTVLAARN--PAKWTMDLHNT--GFLPV--AHPVSLSGMIS-----AEGMN 663  
 Db 1188 --ATILANSNLGANTISGLNHSIDLGNPDLQTRAFMEVLTQILOGTEFTLAEIVL 1245  
 QY 664 AGGYGNL--LMNSEQOGEYGVITAFRLITTLVKGQSGTOSQGLVPCVMFVLKEML-P 720  
 Db 1246 ADRFQVLQVLTMTSDKGEPLIAMAALANVYTTSGMDLAR-----VLVTLFDKHLISF 1299  
 QY 721 SYHKRYSNGVRQIGCLILELIHAILNICHEFTDLHSHTPSIQFLCTSLAYTEAGOT 780  
 Db 1300 LLMNMFREVEVSD-----CMQT--LFRGNSLGSKIMAFCKIY----- 1336  
 QY 781 VINIMGVVDIIDVMAAPRSDGABEGGQQLIKTVKLAFSYTNVNIIRKPPSNVSP 840  
 Db 1337 -----GASYLQMLLEPLIR-----PLDDEEBETCFEV--DPAKLDPTEDI--- 1374  
 QY 841 LEQALSGHAGNNLIIVATKY--IYHKDPALPRLAIQ-----LKLKATVAPMSV 890  
 Db 1375 -EQ-----HRNNLIALTQKVPAIINSSBRLPQLRSMCHCLUYLSKGFPLN----- 1421  
 QY 891 YACGNDAAIRDAFLRLL-----QSKIEDMRKIKWMLLEFITVAVE 931  
 Db 1422 --QNNNGAVGTVIFLFINPAIVSPQELIVDKQVHSSKRGMLSKLQNIANHVE 1478  
 QY 932 TQPLGIELFNLKVDKSDSKERSLGWS--CLHAYELIDISQOQDRWCPLLHRAIA 990  
 Db 1479 FSKQGHKLCFNDPLRDFEAGRRFFTOASCC-----ETVDQTHSM--SFLSDAVL 1529  
 QY 991 FLH-ALW-----QDRDSAMLVLRTPKEMENLTSPFLPGLSPSETSPSILLET 1039

Db 1530 ALHRLMTHQKIGDYSSSRDHKAVGRPFDM-----ATLLAVLGPBEHKPVDSHMF 1584  
 Qy 1040 CALMKIKCEIYVYVVGSLDQSLKD-----TLKKRSI-----EKQFA 1077  
 Db 1585 SS-----YARSSIMSSSTNPEIWKQMEKEEFTKLSNNIYYQAGTSKSG 1633  
 Qy 1078 YMSGVKSLAVVAETEGSSCTSLLEYQMLV-----SARMLLIATTHADIMHLTDS 1130  
 Db 1634 YPVYIYARRYKIGETNG-----DLIYHVIITLKPFCHSPREVIDPHTGSDRFRREF 1689  
 Qy 1131 VVRQQLFDVL--DGTALLVVPASVNCRLRSGSKCTILLILLKQWRELGSV--DEIIGP 1187  
 Db 1690 LQKMFYVLPVYAYENVAHYI-----YNC-----NSWREYTKFPHRIIAP 1730  
 Qy 1188 L-----TELLEGVLQADQO-----LMEKTKAVPSAFI----- 1215  
 Db 1731 LKGRKLLPLESPNKLTDFIDBQOKLPFATLSLDEDLKVSNAIKSHKOTKVAIKXP 1790  
 Qy 1216 TVLQMKEMKASDIPOYSQVLNVCELTQOEVIYALFDQTRHSIALGSATEDKDSMETDGS 1275  
 Db 1791 TALQITSAEKTQVLAHSLVTLNDVYVYASIEVCLVDNQFLTSTINES--GQLSPFHND- 1848  
 Qy 1276 RSRHRDQDQVCLGLHLAK--ELCEVDDEGDSMLQVTRLPILP-----TLTTLLEVSL 1328  
 Db 1849 -----DNIVQAIHIRNMEISQPD-----SYVHOKIRPKQVPPGTLMMALINT 1893  
 Qy 1329 -RMKONLHFEATLHLTLTARTQOGATAVAGAGI--TQSLCPLPSLVYQSTNGAOTPS 1386  
 Db 1894 GSCPNLR--TAAYNLGALATATPD--LKIGQLLETQGLCP-----SNNTTIFKS 1941  
 Qy 1387 ASRSLDAPSWPGVYRLSMSIMEQLKTLRYNFPLEALDFVGHQERTLOCLNAVRVQS 1446  
 Db 1942 VSEKL--ATNEP--HLTFLEESIQGF-----QRTTIELKH----- 1974  
 Qy 1447 LACIEADHTVGFILQSLNFKEMKHFHLPQMLRDIOVNLGYLCAQCSILHSRKLQHYL 1506  
 Db 1975 -LCLE-----YMTPW-----LKNLYKFCNSNDKSKLKVYSQI 2005  
 Qy 1507 ONKKGDLPSVAORVQRPASASAPSSKOPADTEASEQALH--TVQYGL-----LKI 1561  
 Db 2006 LDK-----LINTIDQKEMVPSVQAKINGSIQIPELIDMVDNLFHSITYGGLSPQVEI 2061  
 Qy 1562 LSKTLAAL-----RHFTPDVCOIL-----LDQSL--DLAEY-NFLPALISF 1598  
 Db 2062 MADPAVVALASANYOLVSKVITRICRVMDKCTNPQYQLQHMMDIALIIGRYILMISF 2121  
 Qy 1599 TTFPFDEBVAFSGTILATV-----NVALNMLGELDKKKEPLQVAGLS 1642  
 Db 2122 NNCL--DVATSVYLFHTITFLVCSGSPKRASTHGLVNIHSLCTNP-----S 2171  
 Qy 1643 TOAEGTRTLK-SLIMFTWENCFYLLI-----SOMRYLRDPAVHPRDK--ORMKOEL 1691  
 Db 2172 FSEBAQVLRSLDEFSLPK-FYLLFGISKYKSAAVAFRSCCHPDKMLGNERYQPL 2230  
 Qy 1692 SSELSTL-LSSL 1702  
 Db 2231 PADRERLSLPSL 2242

## RESULT 38

T40186  
 Probable phosphatidylinositol 3-kinase - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C:Accession: T40186

R:Wood, V.; Rajandream, M.A.; Barrall, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.  
 Submitted to the EMBL Data Library, February 1998  
 A:Reference number: Z21910  
 A:Accession: T40186

A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Reidues: 1-2335 <MOO>

A:Cross-references: UNIPROT:O14356; EMBL:Z97992; PIDD:CA10805.1; GSPDB:GNO0067; SPDB:SH

A:Experimental source: strain 972h-, cosmid c30D10  
 C:Genetics:  
 A:Gene: SPDB:SPBC30D10.10c  
 A:Map position: 2  
 C:Superfamily: yeast TOR2 protein

Query Match 1.5%; Score 135.5; DB 2; Length 2335;  
 Best Local Similarity 18.2%; Pred. No.3.7;  
 Matches 279; Conservative 219; Mismatches 526; Indels 507; Gaps 71;

Qy 209 VOCLREQSMIL---EIF---LYAYFEMAPSDILVTKQFKEGFGSGRQTRRLVD 259  
 Db 198 VVCCOREAKVOLQCFNEVLQAEHGLRQSSVEYHGLSLAYKELEPEKSG-----SFR 249  
 Qy 260 ETMDPFYDRIGYSALIVGMDIESLHKCALDREHLOPAQOGLICODMDCLMTEFGD 319  
 Db 250 EHYTEPCD-----LALLRHRD--NSIRRCIVELPLTSEYKPKFGQRIYDSFV--- 299  
 Qy 320 IPHAPVTLAMALLRHTNPEETSVVKKIGTAIQLN--VFOYTRTLQSLASGNDCT 377  
 Db 300 -----YLSHRKQKESLAEALGRIMAVNEMIPYLONILKVIKD----- 342  
 Qy 378 TSTRCM-----CYGGLSFVLSLEHTLGNQODIIDTAGEVADP--SLPELFWGTE 428  
 Db 343 TLAKVREKTOYEKPVFEICIGMLAAAYKL-----ELLEDSSRLGLIFSC 388  
 Qy 429 PTSGAGITLDSVCGMPFHLSP-----LLOLRALVSGSKTAKKYVSFLDKMSFYNELYKH 484  
 Db 389 LSVHLRQALVKAMENIPPLAPTOERILNVSQIITKNEIRI-----NDYTP 438  
 Qy 485 KPHDVISHEDGTLMRQTPRLVPLGGQTNLRIPQGTVGVMLDRAVLRWEXSYSWT 544  
 Db 439 SFNTIYASRPDRQSKSTESITLAL-----ETLG-----TFNFTGYS 475  
 Qy 545 LPTCEIMLHV-----VSTADVQHCQRYKP-----ITDLVHYIS 581  
 Db 476 LISFIOGSVLSYENDNSEIRIAARTCCQVAPADP--CRKTNPLAVESVAEVLKLT 533  
 Qy 582 TDISIADCLPITSRIYML--ORLTIVISPPVVIASVCNCLV----- 624  
 Db 534 --LGIDSDPKIETVSLDERFDRHLAPDN-----HCLPLANDVEFSIREIAII 586  
 Qy 625 --LAARNPAKWTDLRHTGFLPFVAHPVSSLSQMSIABEGMNAAGYGNLMSSEQPGEX 681  
 Db 587 IGRLATVNPAPVMSLRKT-----IIQLLS--DMEVSG-----NSRQKE-- 623  
 Qy 682 GVTIAPRLTLTVKQGLSTQSGVPCVMFVLKMLPBYHKRNVSHGVRREGICLL 741  
 Db 624 -----ESAQLIKLIVSKARTIKPY-----IQ 645  
 Qy 742 ELIHAILNLCHETDLSSHTPSIQFLCTCSLAYTEAGQTVINIMGIVDTIDMWMAAQR 801  
 Db 646 SIHIVILPKRADT---SPGVSSAIIISALGELASVEGSDMVDVDRGSMKI--LVNQQ 700  
 Qy 802 SDGAEQGGQGLIKTVKLAFSVTN-----NVIRLKPSNVVSPLEQALSQH 848  
 Db 701 SSTLKRLASLKCRKLCSRGVYIQPYLDYPRLLGALIGILQSEBPPIRREVLRTIGVL 760  
 Qy 849 GAHGNMILAVLAKYI-----HKHPALPRLAIQLKLKRLATVAPSVVACIG 895  
 Db 761 GA-----LDPTYITTEEVSDDIQSSNNHAGVQISAAQVPSLENTAYMAVATLLIG 812  
 Qy 896 NDAAIRDAFLTLQSLKIED--NRKIMILEFTVAVETOPGLIEFLNLEVXDSGSK 953  
 Db 813 -----ILKDSLSMHHSSVVGAVVHHICQMSKSKTVLP---PQVVPFLQV--MQSLSSSA 864  
 Qy 954 EFSIGMWSCHLAVLELIDSQODRWCPPLHRAALAFHALMODRRDSAMLVLRPKF 1013  
 Db 865 EFYFQQLTTLTSI-----IGBNIRDYVSDIFNLKRVF 896  
 Qy 1014 MENTSLPFLGTLPPSTSEPSILENCALMKIKCEIYVYVVGSLDQSLKDTLKRSIE 1073  
 Db 897 WESTTSLT-----VILE-----LIDALAIALODEFKFYLPQILSCMLKARSLD 940

QY 1074 KRFAYMSGYKSLAVHVAETEGSSCTSLLEYOMLVSAWRMLLIATTHADIMHLDTSVVR 1133  
 DB 941 -----NTSSRSVSYKVLDS-----FVIGSNNEEMHMLPVYII 974  
 QY 1134 RQLEFDVLDGTAKLLVPAVSNCLRLSGMCKTLLILRLQWKRELGSDVDELGPITEL- 1192  
 DB 975 RSFERDT-----FPLG---FRKSALKC---IAQFQSVNFSDBASRIIHLPLVRMLG 1019  
 QY 1193 --EGLVADQOOLMEKTKAV-----PSAFTVLAQ--MKEMKVSDIPQSOVLVAVC--E 1240  
 DB 1020 KSNEDLRA--VIMPTLCALVSQOLGYDSIFLPMNVKVLVSHKISH--PAYELLVRELKGE 1076  
 QY 1241 TLQEEVIALFDQTHSLALGSATEDK--DSMETDCS-----RSRRDQDQGVCLGLHL 1293  
 DB 1077 PLRPDQVYKFKRPSTKPFSS--TQDEVITKLQVDOASLKAAMESSQKLTRO----- 1126  
 QY 1294 AKELCEVEDSDMLQVTRRLPLPTLLTLEVSILRMKQNLHFTETLHLLTLARTQOG 1353  
 DB 1127 -----DWQDWI---RRISI--ELKESPSALRSCTL----- 1154  
 QY 1354 ATAVAGAGITOSICPLISLV-----QLSTNGTRAQTPSASRKSLDAPSMGVRLKMSL 1407  
 DB 1155 -----AGIYHPLARDLFNVVSFLSCWDELTESNKNLVKSIELMNAFN-----ISVEI 1202  
 QY 1408 MEOLLKTLRY-----NFLPEALDFGVYHQR-----TLQCLN-----AVPTVQ 1445  
 DB 1203 LQTLINLAIEYEREDHTLPRIKVISASAKCNVYAKALHYTELOPQVETKEEVSISTIE 1262  
 QY 1446 SLAC-----LEBADHTVGFILQLSNFMKE-----WHPLPOLMR----- 1479  
 DB 1263 SLITINMHLQSDAAG--MLQYTEKHQFSLKETWY---EKLHNMWDLALAEHREBERGD 1318  
 QY 1480 -DIVNLGYLQCACTSLSHSRKMLQHTLQNKQNGCLPSAVQVORPPSAAAPS----- 1534  
 DB 1319 SSFEPINIGKL--RCYVALGMDHLSLQAKA---WVTSQEHREALIPLAAMAAMGLQGW 1373  
 QY 1535 ---SSQPADPTEASEO-----ALHTVOY 1556  
 DB 1374 NLISEYVSANDRPQDKKEFFSAISAVHLQY 1404  
 RESULT 39  
 T42735  
 TBP-interacting protein TBP120 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 11-Jan-2000 #sequence revision 11-Jan-2000 #text change 09-Jul-2004  
 C/Accession: T42735  
 R:Yogorawa, S.; Makino, Y.; Yoshida, T.; Kishimoto, T.; Muramatsu, M.; Tamura, T.  
 Biochem. Biophys. Res. Commun. 229, 612-617, 1996  
 A>Title: Molecular cloning of a novel 120-kDa TBP-interacting protein.  
 A/Reference number: 222253; MIMD:97127450; PMID:8954946  
 A/Accession: T42735  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-1230 <YOG>  
 A/Cross-references: UNIPROT:P97536; EMBL:D87671; NID:g1799569; PTDN:BAAL3432.1; PID:g1795  
 C/GeneticS:  
 A/Gene: tbp120  
 A/Function:  
 A/Description: considered to participate in transcription regulation through the interac  
 Query Match 1.54; Score 135; DB 2; Length 1230;  
 Best Local Similarity 19.28; Pred. No. 1.4;  
 Matches 269; Conservative 191; Mismatches 468; Indels 470; Gaps 68;  
 QY 255 RLHVDFTMDPFDRIQGYFSAIILVEGMDIESLHKCALDRRELHQFADGLICQDMDCIM 314  
 DB 65 QNLAVKCLGPIVSKVKEQVETIV-----DTLCMTMLSDKQQLDSSIG-----LK 111  
 QY 315 LTFQDIPHHAFVLLAMALLRHTLNPETSSVVRKIGGTALQANFOYLTRLLQSLASGN 374  
 DB 112 TVIGELPPAS-----SGSALAAVCKKITGTLSALAKOE 146

QY 375 DCTTACMCVYGLLSFVLTSLEHTLGNODIIDTACEVLAD--PS-----LPELFMGT 427  
 DB 147 D-----VSQLEADIDMADMLSRQGLLVNHPHSLITCLQL---T 185  
 QY 428 EP-----TSGGILDSVCG-----MFPHLSP-----LQLLRALVSG 461  
 DB 186 SPRLAVAKRTIHALGHLVMS--CGNIVFVDILHLLSELSKNDMSMTTRTYIQCIALL--S 242  
 QY 462 KSTAKKYSPDK-----MSFYN-----ELYKHKHVDVSHEDGTLMRQTPKLYPLGQ 512  
 DB 243 RQAGRIRGEYLEKILPLVVKCNDDEL---REYCIQAFSS---FVRCKEYVP--- 292  
 QY 513 TNLRIPOGTQVWLDDRAVLVRMEY-----SYSS-----SW 543  
 DB 293 -----HVSTIINTCLKLTITDPNYVDEDEDENAMADGDDDDGSDDEYSDDDMGW 347  
 QY 544 TLFTCEIEMLLHVSTADVI--QHCRVKPII-----DLVHKVIS----- 581  
 DB 348 KVRRAAKCLDAVASTHEMLPEFYKTVSPALLSRFKEREENVKADVFHAYLSLKQTRP 407  
 QY 582 --TDLSTADCLLPITSKIYMLQRLTVI-----SPRYDVASCVCNCTVLAARPA 631  
 DB 408 VQSWLCPDPMEOGETPLTMOQVPIVYALHKOMEKESVKTQCCFNMTELNVYLFG 467  
 QY 632 KVMTDLHHTGLP-FVAHVPVSSLSQMTSABGMNAGVGNLMMNSEQPOGEVYIAFLRL 690  
 DB 468 AL-----TQHTFVLVGGITFSLNDKSS-----SNLKIDN-----LSCLYV 503  
 QY 691 ITTLVKQQLSTOSQGLVPCVMFVLKEMLPYHMKRYNSGVREIOCLILIELHAILML 750  
 DB 504 ILCHNSFOVHHPHQAVALPVPVACVGD--PPY-----KISEALLVYQQLVKV 549  
 QY 751 CHETDLHSH--TPSLOFLCISLAYTEA-----GQTVINIMIGVDTID 793  
 DB 550 IRPLDQSSDPATFYIKDLFTCTIKRLKADIDQEVKRAISCWQICIMVL----- 600  
 QY 794 MVMMAQPRSGABEQGGQGLIKTVKLAFS--VIMNVIRLKPSPVNV-----SPLEQALSH 848  
 DB 601 -----GDNIGPDLNSTLQIFLERLKNKEITRLTVKATLTIAGSPKIDIRPV 647  
 QY 849 GAHGNMLIAVLAKYIYKHPDPAFLRLAIQLKRLATVAPMSVYACLGND--AAIRDAFL 906  
 DB 648 LGEG---VPIIASFL-RKNGRAL-----KLGTSALDILIKYNSDSITAMIDAVL 694  
 QY 907 TRLOSKI--EDMRIKVILEFLVAVETQGLLEFLINLEVKQSDGSKFSLGMSCLH 964  
 DB 695 DELPPLISESDMHSVQMAISFLTLAKYPS-----SLKISG-----SILN 736  
 QY 965 AVELIDSOQODRWCPPLHRAAIAFLHLMQDRRSAMLVLRKPKFMEV-----LTSP 1020  
 DB 737 ELIGLVS-----PLIQGGLASMLDPFQ---ALVVGTNNLGVWDLRLMLTGP 782  
 QY 1021 LFGTLSPPESTSEPSILETALIMKICLIYYVYVKSLSQSLKDTLKFSIE----- 1073  
 DB 783 VYSGSTALTHKQSYYSIAKCVAALTRACPRGAVVQGFQDYVANGSSTSRILALLSL 842  
 QY 1074 -----KRFAYMSGYKSLAVHVAETEGS--SCTSLLEYQMLVSAMR 1112  
 DB 843 GEVGHNIDLSGOLELKSIVLEAFSPSEBVKSASAYAL---GSISVGNLPEVLPFV----- 895  
 QY 1113 MLIIATTHADIMHLDTSVVRQQLF---DVDGTAKALLVPAVSN--CLRLSGMCKTLL 1166  
 DB 896 -----LQETTSQPRQYLLHSLKEILISSASVAGIKPVENIMALLHGHCECA- 943  
 QY 1167 LLILRLQWKRELGSS--VDEILGPTEILEGVLAQOOLMEKTKAKYFS---AFTVLAQ 1219  
 DB 944 -----EBGRNVVAACLGKLT-----LIDPETLPRLKGYLISSSVARSVVV 987  
 QY 1220 MKEMKVSDIPQSOVLVAVC-----ETLQEEVIALFDQTHSLAL---GSATEDKDSMET 1271  
 DB 988 AVFTTISHDQPIDPLKNCIGDFLKTLEDP-----DLNVRVALVTFNSAAHNPSEL-- 1040

QY 1272 DDCSRSHRDRDQGVCLGLHL-----AKELCEVEDG-----DSMLQVTR-RLEPLP 1318  
DB 1041 -----IRDLDBVLPR-----HLVNETKREKLIREVEMGPKKHIVVDGLDRKAAFEQCMY 1090  
QY 1319 TLLTLEVLNLMKONLHFTETATL--HLLLTARTQOGATAVAGITQSCILPLLSVYL 1376  
DB 1091 TLLDSCIDRLDIFEFILNHVEDGLKDHVDIKM-----LTFELMLVRL 1130  
QY 1377 STNCTACTPSASRSLSAPSWPGVYRSLMSLMEOGLTKRYNFIPEALDPVQVQERTIQ 1436  
DB 1131 ST-----LCPSAVLQRLD-----RLVEPLRATCTTKVXNSVQKEPE-----KODELKR 1174  
QY 1437 CINAARTVQSLACEAD 1454  
DB 1175 --SAMRAVALLTPEAE 1190

## RESULT 40

AC2012  
hypothetical protein a111649 (imported) - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence revision 14-Dec-2001 #text change 09-Jul-2004  
C:Accession: AC2012  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Saeameto, S.; Watanabe, A.; Iriyuchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Ref. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AC2012  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1570 <KUR>  
A:Cross-references: UNIPROT:Q8YWG4; GB:BA000019; PIDN:BA878015.1; PID:917135469; GSPDB:C  
A:Experimental source: strain PCC 7120  
C:Gene: all11649  
C:Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac  
homology; [acyl-carrier-protein] S-malonyltransferase homology  
C:Keywords: carrier protein

Query Match 1.5%; Score 135; DB 2; Length 1570;  
Best Local Similarity 19.6%; Pred. No. 2.1;  
Matches 185; Conservative 116; Mismatches 317; Indels 324; Gaps 44;  
QY 321 PHNAPVLLAMLLHNTLPETSSVKKIGSTALQANVFOYLRLLSLASGNDCTTST 380  
DB 800 PH--PVLISGY--INACLRKQEVGVI-----LPSLRKG----- 828  
QY 381 ACWCVYGLSFLVLSL-ELHTLG---NQODIIDTACEVLADPSLP---ELPFGTEPTSGL 433  
DB 829 -----FGERATLTLSLGLTYLTHGSHINWQSLYPDCQWDLPLYPWQRESYWSQSP-- 881  
QY 434 GIILDSVCGMPHLLSPLLQLLRALVSGKSTAKTVSFLDKMSFYNELYKHKPHDIVISHE 493  
DB 882 -----QFQKAL-PASSLNLVLVAGKTE-----QLTQELSNH-----HQ 913  
QY 494 DGLTMRQTPKLLYPLG-GQTNLALPQGTQGVWMLDDRATLYRWKYS-----YSS 542  
DB 914 LSPBAKQFIPQLQLLTLGKSTAKITOD-----LSNARYIEWQLSPLTVDNKTSBAER 967  
QY 543 WTLFTCEI---EMLLHVSTADVI-----OHQGRKPIIIDLVHKVISTDLSIA 587  
DB 968 WLIFSDNOGLGGLAAYVNBSCILVSSGESYBKSSSHYQ--INF-----HQADFOQLLA 1021  
QY 588 DCLLPITSRIYM-----LQRLTTVTISPVP-----DV 614  
DB 1022 DISPEITKVVYVWELSDSINLNSQPCQSHLYLVQALAKISSKIAPOLMIGTQAQAV 1081  
QY 615 IAGC--VM-----CLYTLAARNPAKWTD----- 636  
DB 1082 TPSCGQINVAQTPLMGQVIALLEYPOLMGLIDFGNODVAATATAIEMTAKTGEDEVAF 1141

QY 637 ---LRHTGFLPFAVHPVSSLSQWIS-AEGMNAAGYNLLMNSHQPOGEYGVTTA----- 686  
DB 1142 RDKRIVARLMPISALPAPQPLISDGSYLIITGGLAL-----GLTLAEWLVQ 1190  
QY 687 -FLRLITTVKQGLQGST-----QSQG---LVPCEVFLKEMLPSEYHKRRYNSHG 731  
DB 1191 GASHLVITSRQGLINQSEKQOKIRALENGATVAVADVSQYQMSQLFAQIOLNSPK 1250  
QY 732 VRBOIGLILIELHA--LNLCHETDLHSHTSLQFLCTCSLAYEAGQTVININGIGV 789  
DB 1251 LR-----GITHAAGVLA-----DCSTAQ-----M 1269  
QY 790 DTIDMWAAPRSDGAEQGGQGLIKITVLAESVTNNVIRLRPPSNVVSPLRQALSQHG 849  
DB 1270 DSETFPKVPQPKYTGAMNLHQ---LVQDSLDFVCE-----SSMSALLSROGLHY 1318  
QY 850 AHCNNLIATLAKTYIKHDPALPRAIQ-----LKRATVAPMSVYAC 893  
DB 1319 AANSPFLDGL---VYHQSLGLPLSLINMGPMAGGMAATQGYEVGLKRMG-IEPLLEPTAA 1374  
QY 894 LGNDAAAIRDAFRLRQSKIEDMRIRKVMLEFLT---VAVETQGLIELFLMLEYKDG 949  
DB 1375 LQVLGGLLSA-----SMQTVAAALNMSAFGKIYAAKGVAFLEALLTQESKDG 1424  
QY 950 DSKERSLGMWSCIHAVLELIDSQODRYWCPLLRRAIA--FLHALMODRRDSAMLV 1007  
DB 1425 NGENPFGKLSAALPHRRPALTLTQVQGV-AQVLGHSGSVPEVEQGFPMGMDSLMSV- 1482  
QY 1008 KTYPKWENTSPFLGTLSPSETSEBSILETCALIMKICLETYYVYKSLDQSLKDTL 1067  
DB 1483 ---QFHSLEALLAVSLPSTLVPECPISGDVVSYLNR---EVFAMQLDADDSAMESQ 1534  
QY 1068 KKFSTIEKRFAYMSGYKSLAVHAETEGSCSTSLFEXOMLV 1109  
DB 1535 ASVIVENTTIAQLRG-----LSTAFETBALMEQRIAEIQLALLS 1570

Search completed: December 30, 2004, 15:14:54  
Job time: 121.789 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2004, 15:00:06 ; Search time 0.191018 Seconds  
(without alignments)  
2083.096 Million cell updates/sec

Title: US-10-719-385-26

Perfect score: 36

Sequence: 1 AGGPGV 6

Scoring table:

BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

1: Issued Patents AA: \*  
2: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/6C\_COMB.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	88.9	15	4	US-09-142-524D-34
2	32	88.9	15	4	US-09-142-524D-35
3	32	88.9	18	2	US-09-017-205-10
4	32	88.9	20	3	US-08-467-023-35
5	32	88.9	20	3	US-08-467-023-36
6	32	88.9	30	3	US-08-467-023-67
7	32	88.9	30	3	US-08-467-023-71
8	32	88.9	32	4	US-09-082-358B-40
9	32	88.9	54	2	US-08-400-159-16
10	32	88.9	54	3	US-08-611-729A-16
11	32	88.9	54	4	US-09-195-524-16
12	32	88.9	60	3	US-08-467-023-62
13	32	88.9	66	1	US-08-450-360-10
14	32	88.9	91	4	US-09-198-452A-726
15	32	88.9	108	4	US-09-513-999C-8111
16	32	88.9	135	4	US-09-252-991A-17866
17	32	88.9	171	4	US-09-252-991A-20639
18	32	88.9	194	4	US-09-252-991A-21193
19	32	88.9	196	2	US-08-481-814A-9
20	32	88.9	226	4	US-09-068-740A-5
21	32	88.9	226	4	US-09-252-991A-27728
22	32	88.9	264	4	US-09-252-991A-20374
23	32	88.9	275	3	US-09-036-987A-7
24	32	88.9	275	3	US-09-370-700-7
25	32	88.9	275	4	US-09-603-207-7
26	32	88.9	290	4	US-09-252-991A-23586
27	32	88.9	365	2	US-08-481-814A-10

28	32	88.9	374	3	US-08-467-023-2	Sequence 2, Appli
29	32	88.9	437	1	US-08-136-119-4	Sequence 4, Appli
30	32	88.9	437	1	US-07-882-711-2	Sequence 2, Appli
31	32	88.9	437	2	US-08-723-415B-13	Sequence 13, Appli
32	32	88.9	437	2	US-08-481-814A-6	Sequence 6, Appli
33	32	88.9	437	2	US-08-462-174-2	Sequence 2, Appli
34	32	88.9	437	3	US-08-801-092-1	Sequence 1, Appli
35	32	88.9	437	3	US-09-189-627A-13	Sequence 13, Appli
36	32	88.9	437	3	US-09-242-737-2	Sequence 2, Appli
37	32	88.9	437	3	US-09-315-113-1	Sequence 1, Appli
38	32	88.9	437	3	US-09-710-861-13	Sequence 13, Appli
39	32	88.9	447	1	US-08-450-360-4	Sequence 4, Appli
40	32	88.9	476	2	US-08-139-937-14	Sequence 14, Appli
41	32	88.9	476	5	PCT-US93-11310-14	Sequence 14, Appli
42	32	88.9	486	1	US-08-450-360-2	Sequence 2, Appli
43	32	88.9	744	2	US-08-462-080B-2	Sequence 2, Appli
44	32	88.9	744	3	US-08-462-090-2	Sequence 2, Appli
45	32	88.9	744	3	US-08-463-461-2	Sequence 2, Appli
46	32	88.9	1010	3	US-08-882-046-7	Sequence 7, Appli
47	32	88.9	1010	4	US-09-556-047-7	Sequence 7, Appli
48	32	88.9	1036	4	US-09-068-740A-6	Sequence 6, Appli
49	32	88.9	1067	4	US-09-579-536C-18	Sequence 18, Appli
50	32	88.9	1160	3	US-08-808-599A-24	Sequence 24, Appli
51	32	88.9	1187	3	US-09-068-740A-7	Sequence 7, Appli
52	32	88.9	1193	2	US-08-400-159-10	Sequence 10, Appli
53	32	88.9	1193	3	US-08-611-729A-10	Sequence 10, Appli
54	32	88.9	1193	4	US-09-195-524-10	Sequence 10, Appli
55	32	88.9	1208	4	US-09-199-865-1	Sequence 1, Appli
56	32	88.9	1218	2	US-08-400-159-6	Sequence 6, Appli
57	32	88.9	1218	3	US-08-611-729A-6	Sequence 6, Appli
58	32	88.9	1218	3	US-08-882-046-2	Sequence 2, Appli
59	32	88.9	1218	3	US-09-214-278-7	Sequence 7, Appli
60	32	88.9	1218	3	US-09-068-740A-11	Sequence 11, Appli
61	32	88.9	1218	4	US-09-855-722-7	Sequence 7, Appli
62	32	88.9	1218	4	US-09-566-047-2	Sequence 2, Appli
63	32	88.9	1218	4	US-09-917-254-85	Sequence 85, Appli
64	32	88.9	1218	4	US-09-195-524-6	Sequence 6, Appli
65	32	88.9	1218	4	US-09-579-536C-1	Sequence 1, Appli
66	32	88.9	1219	3	US-08-882-046-5	Sequence 5, Appli
67	32	88.9	1219	4	US-09-566-047-5	Sequence 5, Appli
68	32	88.9	1404	2	US-08-400-159-2	Sequence 2, Appli
69	32	88.9	1404	2	US-08-611-729A-2	Sequence 2, Appli
70	32	88.9	1404	2	US-09-195-524-2	Sequence 2, Appli
71	31	86.1	27	3	US-09-219-983A-35	Sequence 35, Appli
72	31	86.1	780	4	US-10-148-806-35	Sequence 35, Appli
73	30	83.3	92	4	US-09-513-999C-6101	Sequence 6101, Ap
74	29	80.6	61	4	US-09-621-976-7028	Sequence 7028, Ap
75	29	80.6	67	4	US-09-621-976-6596	Sequence 6596, Ap
76	29	80.6	69	4	US-09-621-976-6022	Sequence 6022, Ap
77	29	80.6	103	4	US-09-513-999C-7000	Sequence 7000, Ap
78	29	80.6	106	3	US-09-199-637A-233	Sequence 233, Ap
79	29	80.6	126	4	US-09-205-258-1076	Sequence 1076, Ap
80	29	80.6	143	4	US-09-252-991A-17861	Sequence 17861, A
81	29	80.6	147	4	US-09-252-991A-31725	Sequence 31725, A
82	29	80.6	235	4	US-09-252-991A-24522	Sequence 24522, A
83	29	80.6	244	4	US-09-769-482-112	Sequence 12, Appli
84	29	80.6	249	4	US-09-311-784A-16	Sequence 16, Appli
85	29	80.6	254	4	US-09-270-767-11730	Sequence 1430, A
86	29	80.6	269	4	US-09-311-021-66	Sequence 66, Appli
87	29	80.6	288	4	US-09-091-952A-4	Sequence 4, Appli
88	29	80.6	306	4	US-09-091-952A-3	Sequence 3, Appli
89	29	80.6	323	4	US-09-252-991A-30172	Sequence 30172, A
90	29	80.6	367	3	US-08-467-023-95	Sequence 95, Appli
91	29	80.6	370	3	US-08-467-023-97	Sequence 97, Appli
92	29	80.6	399	4	US-09-270-767-41122	Sequence 41122, A
93	29	80.6	423	2	US-08-955-713-2	Sequence 2, Appli
94	29	80.6	487	4	US-09-535-909-4	Sequence 4, Appli
95	29	80.6	580	4	US-09-252-991A-29629	Sequence 29629, A
96	29	80.6	592	1	US-09-134-000C-5477	Sequence 5477, Ap
97	29	80.6	968	1	US-08-434-730-14	Sequence 14, Appli
98	29	80.6	2710	2	US-08-458-459A-12	Sequence 12, Appli
99	29	80.6	2710	2	US-08-487-826B-12	Sequence 12, Appli
100	29	80.6	2710	3	US-09-210-288-12	Sequence 12, Appli

SEQ ID NO	LENGTH	TYPE	ORGANISM	FEATURE	NAME/KEY	LOCATION	OTHER INFORMATION
101	29	80.6	3060	2	US-08-487-826B-14	Sequence 14, Appl	
102	28	77.8	8	1	US-08-079-812-34	Sequence 34, Appl	
103	28	77.8	14	1	US-09-741-171-5	Sequence 5, Appl	
104	28	77.8	18	2	US-09-017-205-19	Sequence 19, Appl	
105	28	77.8	19	4	US-09-741-171-4	Sequence 4, Appl	
106	28	77.8	31	4	US-09-084-303B-295	Sequence 295, App	
107	28	77.8	32	4	US-09-084-303B-285	Sequence 285, App	
108	28	77.8	34	4	US-09-749-637A-143	Sequence 143, App	
109	28	77.8	37	1	US-08-149-839B-16	Sequence 16, Appl	
110	28	77.8	37	1	US-08-149-839B-18	Sequence 18, Appl	
111	28	77.8	37	1	US-08-451-568-16	Sequence 16, Appl	
112	28	77.8	37	1	US-08-451-568-18	Sequence 18, Appl	
113	28	77.8	37	1	US-08-451-566-16	Sequence 16, Appl	
114	28	77.8	37	1	US-08-451-566-18	Sequence 18, Appl	
115	28	77.8	37	2	US-08-777-113-16	Sequence 16, Appl	
116	28	77.8	37	2	US-08-777-113-18	Sequence 18, Appl	
117	28	77.8	37	4	US-09-298-574-16	Sequence 16, Appl	
118	28	77.8	37	4	US-09-298-574-18	Sequence 18, Appl	
119	28	77.8	40	3	US-08-851-843A-213	Sequence 213, App	
120	28	77.8	40	3	US-08-851-843A-213	Sequence 213, App	
121	28	77.8	40	3	US-08-854-050-213	Sequence 213, App	
122	28	77.8	40	3	US-09-430-333-213	Sequence 213, App	
123	28	77.8	40	4	US-09-402-181B-332	Sequence 332, App	
124	28	77.8	40	4	US-09-721-456-332	Sequence 332, App	
125	28	77.8	40	4	US-09-721-456-332	Sequence 332, App	
126	28	77.8	46	3	US-09-391-799-5	Sequence 5, Appl	
127	28	77.8	54	4	US-09-149-476-573	Sequence 573, App	
128	28	77.8	54	4	US-08-676-125A-40	Sequence 40, Appl	
129	28	77.8	58	1	US-08-086-328C-31	Sequence 31, Appl	
130	28	77.8	58	1	US-08-086-328C-31	Sequence 31, Appl	
131	28	77.8	58	2	US-09-136-012A-40	Sequence 40, Appl	
132	28	77.8	58	2	US-08-676-124-42	Sequence 42, Appl	
133	28	77.8	58	3	US-08-676-124-43	Sequence 43, Appl	
134	28	77.8	58	3	US-08-676-124-44	Sequence 44, Appl	
135	28	77.8	58	3	US-08-676-124-98	Sequence 98, Appl	
136	28	77.8	58	3	US-08-676-124-100	Sequence 100, App	
137	28	77.8	58	3	US-08-676-124-106	Sequence 106, App	
138	28	77.8	58	3	US-08-676-124-109	Sequence 109, App	
139	28	77.8	58	3	US-08-676-124-111	Sequence 111, App	
140	28	77.8	58	3	US-09-414-878-42	Sequence 42, Appl	
141	28	77.8	58	3	US-09-414-878-43	Sequence 43, Appl	
142	28	77.8	58	3	US-09-414-878-44	Sequence 44, Appl	
143	28	77.8	58	3	US-09-414-878-98	Sequence 98, Appl	
144	28	77.8	58	3	US-09-414-878-100	Sequence 100, App	
145	28	77.8	58	3	US-09-414-878-106	Sequence 106, App	
146	28	77.8	58	3	US-09-414-878-109	Sequence 109, App	
147	28	77.8	58	3	US-09-414-878-111	Sequence 111, App	
148	28	77.8	58	3	US-09-240-136-42	Sequence 42, Appl	
149	28	77.8	58	3	US-09-240-136-43	Sequence 43, Appl	
150	28	77.8	58	3	US-09-240-136-44	Sequence 44, Appl	

## ALIGNMENTS

RESULT 1  
US-09-142-524D-34  
Sequence 34, Application US/09142524D  
Patent No. 671976  
GENERAL INFORMATION:  
APPLICANT: Sone, Toshio  
APPLICANT: Kume, Akinori  
APPLICANT: Dairiki, Kazuo  
APPLICANT: Iwama, Akiho  
APPLICANT: Kuno, Kohsuke  
TITLE OF INVENTION: Peptide-Based Immunotherapeutic Agent for Treating Allergic Disease  
FILE REFERENCE: SPO-103  
CURRENT APPLICATION NUMBER: US/09/142,524D  
CURRENT FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: PCT/JP97/00740  
PRIOR FILING DATE: 1997-03-10  
NUMBER OF SEQ ID NOS: 174  
SOFTWARE: PatentIn version 3.1

SEQ ID NO 34  
LENGTH: 15  
TYPE: PRT  
ORGANISM: Cryptosporidia japonica  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (1)-(15)  
OTHER INFORMATION: Cryj1 peptide, Figure 1, Row 120  
US-09-142-524D-34

Query Match  
Best Local Similarity: 88.9%; Score 32; DB 4; Length 15;  
Best Local Similarity: 100.0%; Pred. No. 47;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
DB 4 GGPCV 8

RESULT 2  
US-09-142-524D-35  
Sequence 35, Application US/09142524D  
Patent No. 671976  
GENERAL INFORMATION:  
APPLICANT: Sone, Toshio  
APPLICANT: Kume, Akinori  
APPLICANT: Dairiki, Kazuo  
APPLICANT: Iwama, Akiho  
APPLICANT: Kuno, Kohsuke  
TITLE OF INVENTION: Peptide-Based Immunotherapeutic Agent for Treating Allergic Disease  
FILE REFERENCE: SPO-103  
CURRENT APPLICATION NUMBER: US/09/142,524D  
CURRENT FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: PCT/JP97/00740  
PRIOR FILING DATE: 1997-03-10  
NUMBER OF SEQ ID NOS: 174  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 35  
LENGTH: 15  
TYPE: PRT  
ORGANISM: Cryptosporidia japonica  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (1)-(15)  
OTHER INFORMATION: Cryj1 peptide, Figure 1, Row 21  
US-09-142-524D-35

Query Match  
Best Local Similarity: 88.9%; Score 32; DB 4; Length 15;  
Best Local Similarity: 100.0%; Pred. No. 47;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3  
US-09-017-205-20  
Sequence 20, Application US/09017205  
Patent No. 5965357  
GENERAL INFORMATION:  
APPLICANT: Marsden, Howard S  
TITLE OF INVENTION: PEPTIDE STRUCTURES AND THEIR USE IN  
DIAGNOSIS OF HERPES SIMPLEX VIRUS TYPE 2  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon & Vanderye PC  
STREET: 8th Floor, 1100 No. 5965357th Glebe Road  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22201-4714  
COMPUTER READABLE FORM:



MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/017,205  
FILING DATE: 02-FEB-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mitchard, Leonard C  
REGISTRATION NUMBER: 29,009  
REFERENCE/DOCKET NUMBER: 604-436  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)816-4000  
TELEFAX: (703)816-4100  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide from HSV-2 glycoprotein G  
FRAGMENT TYPE: internal  
US-09-017-205-20

Query Match 88.9%; Score 32; DB 2; Length 18;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
|||||  
Db 5 GPCV 9

RESULT 4  
US-08-467-023-35  
Sequence 35, Application US/08467023  
Patent No. 6090386  
GENERAL INFORMATION:  
APPLICANT: Griffith, Irwin J.;  
APPLICANT: Pollock, Joanne;  
APPLICANT: Bond, Julian F.;  
APPLICANT: Garman, Richard D;  
APPLICANT: Kuo, Mei-Chang;  
APPLICANT: Yeung, Siu-mei H.;  
APPLICANT: Brauer, Andrew;  
APPLICANT: Exley, Mark A.;  
APPLICANT: Powers, Steven P.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
TITLE OF INVENTION: Japanese Cedar Pollen  
NUMBER OF SEQUENCES: 261  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
STREET: 610 Lincoln St  
CITY: Waltham  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,023  
FILING DATE: June 6, 1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/350,225  
FILING DATE: December 6, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane E. Remillard  
REGISTRATION NUMBER: 38,872  
REFERENCE/DOCKET NUMBER: 025.6 USD2 (IM1-028CPD2)

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-467-023-35

Query Match 88.9%; Score 32; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
|||||  
Db 14 GPCV 18

RESULT 5  
US-08-467-023-36  
Sequence 36, Application US/08467023  
Patent No. 6090386  
GENERAL INFORMATION:  
APPLICANT: Griffith, Irwin J.;  
APPLICANT: Pollock, Joanne;  
APPLICANT: Bond, Julian F.;  
APPLICANT: Garman, Richard D;  
APPLICANT: Kuo, Mei-Chang;  
APPLICANT: Yeung, Siu-mei H.;  
APPLICANT: Brauer, Andrew;  
APPLICANT: Exley, Mark A.;  
APPLICANT: Powers, Steven P.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
TITLE OF INVENTION: Japanese Cedar Pollen  
NUMBER OF SEQUENCES: 261  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
STREET: 610 Lincoln St  
CITY: Waltham  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,023  
FILING DATE: June 6, 1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/350,225  
FILING DATE: December 6, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane E. Remillard  
REGISTRATION NUMBER: 38,872  
REFERENCE/DOCKET NUMBER: 025.6 USD2 (IM1-028CPD2)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-467-023-36

Query Match 88.9%; Score 32; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
|||||  
Db 4 GPCV 8

## RESULT 6

US-08-467-023-67  
; Sequence 67, Application US/08467023  
; Patent No. 6090386  
; GENERAL INFORMATION:  
; APPLICANT: Griffee, Irwin J.;  
; APPLICANT: Pollock, Joanne;  
; APPLICANT: Bond, Julian F.;  
; APPLICANT: Garman, Richard D;  
; APPLICANT: Kuo, Mei-Chang;  
; APPLICANT: Yeung, Siu-mei H.;  
; APPLICANT: Brauer, Andrew;  
; APPLICANT: Exley, Mark A.;  
; APPLICANT: Powers, Steven P.  
; TITLE OF INVENTION: Allergenic Proteins And Peptides From  
; TITLE OF INVENTION: Japanese Cedar Pollen  
; NUMBER OF SEQUENCES: 261  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
; STREET: 610 Lincoln St  
; CITY: Waltham  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,023  
; FILING DATE: June 6, 1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/350,225  
; FILING DATE: December 6, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jane E. Remillard  
; REGISTRATION NUMBER: 38,872  
; REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 67:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
; US-08-467-023-67

Query Match 88.9%; Score 32; DB 3; Length 30;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
|||||  
Db 24 GPCV 28

RESULT 7  
US-08-467-023-71  
; Sequence 71, Application US/08467023

; Patent No. 6090386  
; GENERAL INFORMATION:  
; APPLICANT: Griffee, Irwin J.;  
; APPLICANT: Pollock, Joanne;  
; APPLICANT: Bond, Julian F.;  
; APPLICANT: Garman, Richard D;  
; APPLICANT: Kuo, Mei-Chang;  
; APPLICANT: Yeung, Siu-mei H.;  
; APPLICANT: Brauer, Andrew;  
; APPLICANT: Exley, Mark A.;  
; APPLICANT: Powers, Steven P.  
; TITLE OF INVENTION: Allergenic Proteins And Peptides From  
; TITLE OF INVENTION: Japanese Cedar Pollen  
; NUMBER OF SEQUENCES: 261  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
; STREET: 610 Lincoln St  
; CITY: Waltham  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,023  
; FILING DATE: June 6, 1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/350,225  
; FILING DATE: December 6, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jane E. Remillard  
; REGISTRATION NUMBER: 38,872  
; REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 71:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
; US-08-467-023-71

Query Match 88.9%; Score 32; DB 3; Length 30;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
|||||  
Db 24 GPCV 28

RESULT 8  
US-09-082-358B-40  
; Sequence 40, Application US/09082358B  
; Patent No. 6469153  
; GENERAL INFORMATION:  
; APPLICANT: Goff, Stephen P.  
; APPLICANT: Li, Xingqiang  
; TITLE OF INVENTION: EIP-1, EIP-3 GENES, ENVELOPE-INTERACTING PROTEINS,  
; FILE REFERENCE: 0575/54804  
; CURRENT APPLICATION NUMBER: US/09/082,358B  
; CURRENT FILING DATE: 1998-05-20  
; NUMBER OF SEQ ID NOS: 106  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 40

LENGTH: 32  
TYPE: PRT  
ORGANISM: murine  
US-09-082-358B-40

Query Match 88.9%; Score 32; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 17 AGGPC 21

RESULT 9  
US-08-400-159-16  
Sequence 16, Application US/08400159  
Patent No. 5869282  
GENERAL INFORMATION:  
APPLICANT: Ish-Horowitz, David  
APPLICANT: Henrique, Domingos M.P.  
APPLICANT: Lewis, Julian H.  
APPLICANT: Myat, Anna M.  
APPLICANT: Fleming, Robert J.  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Mann, Robert S.  
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/400.159  
FILING DATE: 07-MAR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-029  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-400-159-16

Query Match 88.9%; Score 32; DB 2; Length 54;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 5 AGGPC 9

RESULT 10  
US-08-611-729A-16

Sequence 16, Application US/08611729A  
Patent No. 6004924  
GENERAL INFORMATION:  
APPLICANT: Ish-Horowitz, David  
APPLICANT: Henrique, Domingos M.P.  
APPLICANT: Lewis, Julian H.  
APPLICANT: Myat, Anna M.  
APPLICANT: Fleming, Robert J.  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Mann, Robert S.  
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611.729A  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-037  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-611-729A-16

Query Match 88.9%; Score 32; DB 3; Length 54;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 5 AGGPC 9

RESULT 11  
US-09-195-524-16  
Sequence 16, Application US/09195524  
Patent No. 6703489  
GENERAL INFORMATION:  
APPLICANT: Ish-Horowitz, David  
APPLICANT: Henrique, Domingos M.P.  
APPLICANT: Lewis, Julian H.  
APPLICANT: Myat, Anna M.  
APPLICANT: Fleming, Robert J.  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Mann, Robert S.  
APPLICANT: Gray, Grace E.  
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/195,524  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/611,729  
FILING DATE: 06-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-037  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-09-195-524-16

Query Match 88.9%; Score 32; DB 4; Length 54;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPC 5  
Db 5 AGPC 9

RESULT 12  
US-08-467-023-62  
Sequence 62, Application US/08467023  
Patent No. 6090386  
GENERAL INFORMATION:  
APPLICANT: Griffith, Irwin J.;  
APPLICANT: Pollock, Joanne;  
APPLICANT: Bond, Julian F.;  
APPLICANT: Garman, Richard D;  
APPLICANT: Kuo, Mei-Chang;  
APPLICANT: Yeung, Siu-mei H.;  
APPLICANT: Brauer, Andrew;  
APPLICANT: Ekley, Mark A.;  
APPLICANT: Powers, Steven P.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
NUMBER OF SEQUENCES: 261  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
STREET: 610 Lincoln St  
CITY: Walcham  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,023

FILING DATE: June 6, 1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/350,225  
FILING DATE: December 6, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane E. Remillard  
REGISTRATION NUMBER: 38,872  
REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 62:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 60 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-467-023-62

Query Match 88.9%; Score 32; DB 3; Length 60;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GPCV 6  
Db 44 GPCV 48

RESULT 13  
US-08-450-360-10  
Sequence 10, Application US/08450360  
Patent No. 5656457  
GENERAL INFORMATION:  
APPLICANT: Parkes, Deborah Lynn  
APPLICANT: Coates, Stephen Ralph  
TITLE OF INVENTION: Herpes Simplex Virus Type 2-Glycoprotein G  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,360  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/129,021  
FILING DATE: 29-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0012.2A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
GENERAL INFORMATION:  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-08-450-360-10

## Query Match

88.9%; Score 32; DB 1; Length 66;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 2 AGPCV 6  
|||||

Db 18 AGPCV 22

RESULT 14  
US-09-198-452A-726

Sequence 726, Application US/09198452A

Patent No. 6559294

GENERAL INFORMATION:

APPLICANT: Griffiths, R.

TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention

TITLE OF INVENTION: and treatment of infection

FILE REFERENCE: 9710-003-999

CURRENT APPLICATION NUMBER: US/09/198,452A

CURRENT FILING DATE: 1998-11-24

NUMBER OF SEQ ID NOS: 6849

SEQ ID NO 726

LENGTH: 91

TYPE: PRT

ORGANISM: Chlamydia pneumoniae

US-09-198-452A-726

## Query Match

88.9%; Score 32; DB 4; Length 91;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 1 AGPC 5  
|||||

Db 78 AGPC 82

RESULT 15  
US-09-513-999C-8111

Sequence 811, Application US/09513999C

Patent No. 6783961

GENERAL INFORMATION:

APPLICANT: Dumas Mline Edwards, J.B.

APPLICANT: Duclet, A.

TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

FILE REFERENCE: 59, US2, REG

CURRENT APPLICATION NUMBER: US/09/513,999C

CURRENT FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/122,487

PRIOR FILING DATE: 1999-02-26

NUMBER OF SEQ ID NOS: 36681

SOFTWARE: Patent.pm

SEQ ID NO 811

LENGTH: 108

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: UNSURE

LOCATION: 6

OTHER INFORMATION: Xaa=Met or Arg

FEATURE:

NAME/KEY: UNSURE

LOCATION: 89

OTHER INFORMATION: Xaa=Ileu or Val

US-09-513-999C-8111

## Query Match

88.9%; Score 32; DB 4; Length 108;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 1 AGPC 5  
|||||

Db 24 AGPC 28

RESULT 16  
US-09-252-991A-17866

Sequence 17866, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 17866

LENGTH: 135

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-17866

## Query Match

88.9%; Score 32; DB 4; Length 135;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 1 AGPC 5  
|||||

Db 72 AGPC 76

RESULT 17  
US-09-252-991A-20639

Sequence 20639, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 20639

LENGTH: 171

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-20639

## Query Match

88.9%; Score 32; DB 4; Length 171;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 1 AGPC 5  
|||||

Db 115 AGPC 119

RESULT 18

US-09-252-991A-21199  
; Sequence 21199, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21199  
; LENGTH: 194  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21199

Query Match 88.9%; Score 32; DB 4; Length 194;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 AGPCV 6  
DB 14 AGPCV 18

RESULT 19  
US-08-481-814A-9

; Sequence 9, Application US/08481814A  
; Patent No. 5869040  
; GENERAL INFORMATION:  
; APPLICANT: Heu, Yen-Ming  
; TITLE OF INVENTION: GENE THERAPY METHODS AND COMPOSITIONS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Biogen, Inc.  
; STREET: 14 Cambridge Center  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: United States of America  
; ZIP: 02142  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,814A  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kaplan, Warren A  
; REFERENCE/DOCKET NUMBER: A001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-679-2000  
; TELEFAX: 617-679-2838  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 196 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-814A-9

Query Match 88.9%; Score 32; DB 2; Length 196;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AGPC 5  
DB 8 AGPC 12

RESULT 20  
US-09-068-740A-5

; Sequence 5, Application US/09068740A  
; Patent No. 6337387  
; GENERAL INFORMATION:  
; APPLICANT: SAKANO, SEIJI  
; APPLICANT: ITOH, AKIRA  
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE  
; FILE REFERENCE: KP-8447  
; CURRENT APPLICATION NUMBER: US/09/068,740A  
; CURRENT FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: JP 7-239611  
; PRIOR FILING DATE: 1995-11-17  
; PRIOR APPLICATION NUMBER: JP 7-311811  
; PRIOR FILING DATE: 1995-11-30  
; PRIOR APPLICATION NUMBER: PCT/JP96/03356  
; PRIOR FILING DATE: 1996-11-15  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 198  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-068-740A-5

Query Match 88.9%; Score 32; DB 3; Length 198;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPC 5  
DB 57 AGPC 61

RESULT 21  
US-09-252-991A-27728

; Sequence 27728, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27728  
; LENGTH: 226  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27728

Query Match 88.9%; Score 32; DB 4; Length 226;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPC 5  
DB 104 AGPC 108

RESULT 22

US-09-252-991A-20374  
; Sequence 20374, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 20374  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20374

Query Match 88.9%; Score 32; DB 4; Length 264;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 AGPC 5  
DB 198 AGPC 202

RESULT 23  
US-09-036-987A-7  
; Sequence 7, Application US/09036987A  
; Patent No. 6143526  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Merlo, Donald J.  
; APPLICANT: Treadway, Patli J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; TITLE OF INVENTION: Production  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dow AgroSciences LLC Patent Department  
; STREET: 9330 Zionsville Road  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/036,987A  
; FILING DATE: 09-MAR-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stuart, Donald R  
; REGISTRATION NUMBER: 28,479  
; REFERENCE/DOCKET NUMBER: 50,608  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317)337-4816  
; TELEFAX: (317)337-4847  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 275 amino acids  
; TYPE: amino acid

TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-036-987A-7

Query Match 88.9%; Score 32; DB 3; Length 275;  
Best Local Similarity 100.0%; Pred. No. 5.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 AGPC 5  
DB 27 AGPC 31

RESULT 24  
US-09-370-700-7  
; Sequence 7, Application US/09370700  
; Patent No. 6274350  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Treadway, Patli J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50489 Div1  
; CURRENT APPLICATION NUMBER: US/09/370,700  
; CURRENT FILING DATE: 1999-08-09  
; EARLIER APPLICATION NUMBER: US 09/36987  
; EARLIER FILING DATE: 1998-03-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 275  
; TYPE: PRT  
; ORGANISM: Saccharopolyspora spinosa  
US-09-370-700-7

Query Match 88.9%; Score 32; DB 3; Length 275;  
Best Local Similarity 100.0%; Pred. No. 5.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 AGPC 5  
DB 27 AGPC 31

RESULT 25  
US-09-603-207-7  
; Sequence 7, Application US/09603207B  
; Patent No. 6521406  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Treadway, Patli J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50489 Div1  
; CURRENT APPLICATION NUMBER: US/09/603,207B  
; CURRENT FILING DATE: 2000-06-23  
; EARLIER APPLICATION NUMBER: 09/370,700  
; EARLIER FILING DATE: 1998-03-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 275  
; TYPE: PRT  
; ORGANISM: Saccharopolyspora spinosa  
US-09-603-207-7



Query Match 88.9%; Score 32; DB 4; Length 275;  
Best Local Similarity 100.0%; Pred. No. 5.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
|||  
Db 27 AGGPC 31

RESULT 26  
US-09-252-991A-23586  
; Sequence 23586, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23586  
; LENGTH: 290  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23586

Query Match 88.9%; Score 32; DB 4; Length 290;  
Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
|||  
Db 18 AGGPC 22

RESULT 27  
US-08-481-814A-10  
; Sequence 10, Application US/08481814A  
; Patent No. 5869040  
; GENERAL INFORMATION:  
; APPLICANT: Hsu, Yen-Ming  
; TITLE OF INVENTION: GENE THERAPY METHODS AND COMPOSITIONS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Biogen, Inc.  
; STREET: 14 Cambridge Center  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: United States of America  
; ZIP: 02142  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,814A  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kaplan, Warren A  
; REFERENCE/DOCKET NUMBER: A001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-679-2000  
; TELEFAX: 617-679-2838  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 365 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-814A-10

Query Match 88.9%; Score 32; DB 2; Length 365;  
Best Local Similarity 100.0%; Pred. No. 7.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
|||  
Db 8 AGGPC 12

RESULT 28  
US-08-467-023-2  
; Sequence 2, Application US/08467023  
; Patent No. 6090386  
; GENERAL INFORMATION:  
; APPLICANT: Griffler, Irwin J.;  
; APPLICANT: Pollock, Joanne;  
; APPLICANT: Bond, Julian F.;  
; APPLICANT: Garman, Richard D;  
; APPLICANT: Kuo, Mei-Chang;  
; APPLICANT: Yeung, Siu-mei H.;  
; APPLICANT: Brauer, Andrew;  
; APPLICANT: Exley, Mark A.;  
; APPLICANT: Powers, Steven P.  
; TITLE OF INVENTION: Allergenic Proteins And Peptides From  
; TITLE OF INVENTION: Japanese Cedar Pollen  
; NUMBER OF SEQUENCES: 261  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
; STREET: 610 Lincoln St  
; CITY: Waltham  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,023  
; FILING DATE: June 6, 1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/350,225  
; FILING DATE: December 6, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jane E. Remillard  
; REGISTRATION NUMBER: 38,872  
; REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 374 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-467-023-2

Query Match 88.9%; Score 32; DB 3; Length 374;  
Best Local Similarity 100.0%; Pred. No. 7.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGPCV 6  
|||  
Db 125 GGPCV 129

## RESULT 29

US-08-136-119-4  
Sequence 4, Application US/08136119  
Patent No. 5473056  
GENERAL INFORMATION:  
APPLICANT: Heimbrock, David C.  
APPLICANT: Hoyle, Mona I.  
APPLICANT: Olfie, Allen I.  
TITLE OF INVENTION: E 2F-2, A NOVEL MAMMALIAN TRANSCRIPTION  
TITLE OF INVENTION: FACTOR  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Muthard  
STREET: P.O. Box 2000, 126 Lincoln Avenue  
CITY: Rahway  
STATE: N.J.  
COUNTRY: USA  
ZIP: 07065-0907  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/136,119  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Muthard, David A.  
REGISTRATION NUMBER: 35,297  
REFERENCE/DOCKET NUMBER: 19042  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)594-3903  
TELEFAX: (908)594-4720  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-136-119-4  
Query Match 88.9%; Score 32; DB 1; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGGPC 5  
|||  
Db 8 AGGPC 12

## RESULT 30

US-07-882-711-2  
Sequence 2, Application US/07882711  
Patent No. 5759803  
GENERAL INFORMATION:  
APPLICANT: Kaelin Jr., William G.  
APPLICANT: Flemington, Erik  
APPLICANT: Sellers, William  
APPLICANT: Decaprio, James A.  
APPLICANT: Livingston, David M.  
TITLE OF INVENTION: Retinoblastoma Associated Protein-1 cDNA  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Choate, Hall & Stewart  
STREET: Exchange Place, 53 State Street  
CITY: Boston

STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/882,711  
FILING DATE: 19920513  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Kennedy, Bill  
REGISTRATION NUMBER: 33,407  
REFERENCE/DOCKET NUMBER: DPCI#236  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-5020  
TELEFAX: (617) 227-7566  
TELEX: 289374  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
US-07-882-711-2

## Query Match

88.9%; Score 32; DB 1; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
|||  
Db 8 AGGPC 12

## RESULT 31

US-08-723-415B-13  
Sequence 13, Application US/08723415B  
Patent No. 5859199  
GENERAL INFORMATION:  
APPLICANT: Lathague, Nicholas B.  
APPLICANT: delaluna, Susana  
TITLE OF INVENTION: TRANSCRIPTION FACTOR DP-3 AND ISOFORMS  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHYE P.C.  
STREET: 1100 NO. 5859199th Gleebe Rd. 8th floor  
CITY: Arlington  
STATE: VA  
COUNTRY: USA  
ZIP: 22201-4741  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/723,415B  
FILING DATE: 30-SEP-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9610195.1  
FILING DATE: 15-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Crawford, Arthur R.  
REGISTRATION NUMBER: 25,327  
REFERENCE/DOCKET NUMBER: 117-220  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-816-4000

TELEFAX: 703-816-4100  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-723-415B-13

Query Match 88.9%; Score 32; DB 2; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 8 AGGPC 12

## RESULT 32

US-08-481-814A-6  
Sequence 6, Application US/08481814A  
Patent No. 5869040  
GENERAL INFORMATION:  
APPLICANT: Hsu, Yen-Ming  
TITLE OF INVENTION: GENE THERAPY METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Biogen, Inc.  
STREET: 14 Cambridge Center  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: United States of America  
ZIP: 02142  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481.814A  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Kaplan, Warren A  
REFERENCE/DOCKET NUMBER: A001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-679-2000  
TELEFAX: 617-679-2838  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
US-08-481-814A-6

Query Match 88.9%; Score 32; DB 2; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 8 AGGPC 12

RESULT 33  
US-08-462-174-2  
Sequence 2, Application US/08462174  
Patent No. 5981723  
GENERAL INFORMATION:

APPLICANT: Kaelin Jr., William G.  
APPLICANT: Flemington, Erik  
APPLICANT: Sellers, William  
APPLICANT: Decaprio, James A.  
APPLICANT: Livingston, David M.  
TITLE OF INVENTION: Retinoblastoma Associated Protein-1 cDNA  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Choate, Hall & Stewart  
STREET: Exchange Place, 53 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,174  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,711  
FILING DATE: 13-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Kennedy, Bill  
REGISTRATION NUMBER: 33,407  
REFERENCE/DOCKET NUMBER: DPCI236  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-5020  
TELEFAX: (617) 227-7566  
TELEX: 289374

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
US-08-462-174-2

Query Match 88.9%; Score 32; DB 2; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 8 AGGPC 12

RESULT 34  
US-08-801-092-1  
Sequence 1, Application US/08801092  
Patent No. 6074850  
GENERAL INFORMATION:

APPLICANT: Antelman, Douglas  
APPLICANT: Gregory, Richard J.  
APPLICANT: Wile, Kenneth N.  
TITLE OF INVENTION: Tissue Specific Expression of  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/801,092  
FILING DATE: 14-FEB-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/751,517  
FILING DATE: 15-NOV-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Filtes, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 016930-001020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 703-576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
US-08-801-092-1

Query Match 88.9%; Score 32; DB 3; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 8 AGGPC 12

RESULT 35  
US-09-189-627A-13  
Sequence 13, Application US/09189627A  
Patent No. 6159691  
GENERAL INFORMATION:  
APPLICANT: La Thangue, Nicholas  
TITLE OF INVENTION: TRANSCRIPTION FACTOR DP-3 AND ISOFORMS THEREOF  
FILE REFERENCE: 620-54  
CURRENT APPLICATION NUMBER: US/09/189,627A  
CURRENT FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: 08/723,415  
PRIOR FILING DATE: 1996-09-30  
PRIOR APPLICATION NUMBER: GB 9610195  
PRIOR FILING DATE: 1996-05-15  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 13  
LENGTH: 437  
TYPE: PRT  
ORGANISM: human  
US-09-189-627A-13

Query Match 88.9%; Score 32; DB 3; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 8 AGGPC 12

RESULT 36  
US-09-242-737-2  
Sequence 2, Application US/09242737  
Patent No. 6368609  
GENERAL INFORMATION:  
APPLICANT: BERNARDS, REN (

TITLE OF INVENTION: E2F UBIQUITINATION DOMAIN, AND ASSAYS FOR  
INHIBITORS OF E2F UBIQUITINATION  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHAYE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/242,737  
FILING DATE: 23-Feb-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB97/02293  
FILING DATE: 22-AUG-1997  
APPLICATION NUMBER: GB 961697.9  
FILING DATE: 23-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: WILSON, MARY J.  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 620-63  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-242-737-2

Query Match 88.9%; Score 32; DB 3; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 8 AGGPC 12

RESULT 37  
US-09-315-113-1  
Sequence 1, Application US/09315113  
Patent No. 6379927  
GENERAL INFORMATION:  
APPLICANT: Antelman, Douglas  
Gregory, Richard J.  
Wils, Kenneth N.  
TITLE OF INVENTION: Tissue Specific Expression of  
Retinoblastoma Protein  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/315,113  
FILING DATE: 19-May-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/801,092  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: FITEB, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 016930-001020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 703-576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-315-113-1

Query Match 88.9%; Score 32; DB 3; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
|||  
Db 8 AGGPC 12

RESULT 38  
US-09-710-861-13  
Sequence 13, Application US/09710861  
GENERAL INFORMATION:  
APPLICANT: La Thangue, Nicholas  
TITLE OF INVENTION: TRANSCRIPTION FACTOR DP-3 AND ISOFORMS THEREOF  
FILE REFERENCE: 620-54  
CURRENT FILING DATE: US/09/710,861  
PRIOR FILING DATE: 2000-11-13  
PRIOR APPLICATION NUMBER: US/09/189,627  
PRIOR FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: 08/723,415  
PRIOR FILING DATE: 1996-09-30  
PRIOR APPLICATION NUMBER: GB 9610195  
PRIOR FILING DATE: 1996-05-15  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 13  
LENGTH: 437  
TYPE: PRT  
ORGANISM: human  
US-09-710-861-13

Query Match 88.9%; Score 32; DB 3; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
|||  
Db 8 AGGPC 12

RESULT 39  
US-08-450-360-4  
Sequence 4, Application US/08450360  
Patent No. 5656457  
GENERAL INFORMATION:  
APPLICANT: Parkes, Deborah Lynn  
APPLICANT: Coates, Stephen Ralph

TITLE OF INVENTION: Herpes Simplex Virus Type 2-Glycoprotein G  
TITLE OF INVENTION: Proteins and Polypeptides  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (RPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,360  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/129,021  
FILING DATE: 29-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0012.2A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
GENERAL INFORMATION:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 447 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-450-360-4

Query Match 88.9%; Score 32; DB 1; Length 447;  
Best Local Similarity 100.0%; Pred. No. 8.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
|||  
Db 167 GGPCV 171

RESULT 40  
US-08-139-937-14  
Sequence 14, Application US/08139937  
Patent No. 5821070  
GENERAL INFORMATION:  
APPLICANT: LEE, WEN-HWA  
APPLICANT: SHAN, BEI  
TITLE OF INVENTION: CELLULAR GENES ENCODING  
TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CAMPBELL AND FLORES  
STREET: 4370 LA JOLLA VILLAGE DRIVE  
CITY: SAN DIEGO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/139,937  
FILING DATE: 20-OCT-1993  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/979,156  
 FILING DATE: 20-NOV-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: CAMPBELL, CATRYN  
 REGISTRATION NUMBER: 31,815  
 REFERENCE/DOCKET NUMBER: P-CJ 9370  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619-535-9001  
 TELEFAX: 619-535-8949  
 INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 476 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-139-937-14

Query Match 88.9%; Score 32; DB 2; Length 476;  
 Best Local Similarity: 100.0%; Pred. No. 9e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPC 5  
 |||||  
 Db 47 AGGPC 51

Search completed: December 30, 2004, 15:16:02  
 Job time : 7.19102 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 30, 2004, 15:14:08 : Search time 0.736782 Seconds  
(without alignments)  
2929.449 Million cell updates/sec

Title: US-10-719-385-26

Perfect score: 36

Sequence: 1 AGGRCV 6

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1599051 seqs, 359727711 residues

Total number of hits satisfying chosen parameters: 1599051

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 150 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09C\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10D\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
21: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	36	100.0	6 17 US-10-719-385-26	Sequence 26, App1
2	36	100.0	68 17 US-10-425-115-359821	Sequence 359821, App1
3	36	100.0	339 16 US-10-437-963-157034	Sequence 157034, App1
4	36	100.0	407 16 US-10-408-765A-817	Sequence 817, App1
5	36	100.0	858 14 US-10-177-293-47	Sequence 47, App1
6	36	100.0	1745 17 US-10-719-385-21	Sequence 21, App1
7	36	100.0	1745 17 US-10-370-715B-544	Sequence 544, App1
8	36	100.0	3534 17 US-10-719-385-22	Sequence 22, App1
9	35	97.2	206 15 US-10-424-589-230302	Sequence 230302, App1
10	35	97.2	378 15 US-10-424-589-149625	Sequence 149625, App1
11	35	97.2	409 15 US-10-424-589-279664	Sequence 279664, App1
12	35	97.2	1036 9 US-09-771-161A-255	Sequence 255, App1
13	35	97.2	1036 9 US-09-771-161A-256	Sequence 256, App1

14	35	97.2	1036	14	US-10-318-410-2	Sequence 2, App1
15	35	97.2	1092	17	US-10-737-450-72	Sequence 72, App1
16	33	91.7	82	16	US-10-437-963-201950	Sequence 201950, App1
17	33	91.7	93	16	US-10-437-963-153250	Sequence 153250, App1
18	33	91.7	95	16	US-10-437-963-121278	Sequence 121278, App1
19	33	91.7	98	16	US-10-767-701-511998	Sequence 511998, App1
20	33	91.7	149	16	US-10-437-963-170381	Sequence 170381, App1
21	33	91.7	152	16	US-10-437-963-129834	Sequence 129834, App1
22	33	91.7	338	17	US-10-425-115-235622	Sequence 235622, App1
23	33	91.7	351	16	US-10-767-701-40379	Sequence 40379, App1
24	33	91.7	355	16	US-10-767-701-41600	Sequence 41600, App1
25	33	91.7	371	15	US-10-425-114-52879	Sequence 52879, App1
26	33	91.7	782	16	US-10-437-963-180870	Sequence 180870, App1
27	33	91.7	931	16	US-10-437-963-156470	Sequence 156470, App1
28	33	91.7	1824	16	US-10-437-963-194642	Sequence 194642, App1
29	33	88.9	15	14	US-10-354-240-34	Sequence 34, App1
30	32	88.9	15	14	US-10-354-240-35	Sequence 35, App1
31	32	88.9	38	17	US-10-425-115-341764	Sequence 341764, App1
32	32	88.9	43	16	US-10-437-963-186690	Sequence 186690, App1
33	32	88.9	45	17	US-10-425-115-249558	Sequence 249558, App1
34	32	88.9	48	14	US-10-029-386-31343	Sequence 31343, App1
35	32	88.9	48	17	US-10-425-115-286513	Sequence 286513, App1
36	32	88.9	52	15	US-10-424-589-241100	Sequence 241100, App1
37	32	88.9	54	15	US-10-424-589-181045	Sequence 181045, App1
38	32	88.9	57	15	US-10-424-589-152102	Sequence 152102, App1
39	32	88.9	57	17	US-10-425-115-336645	Sequence 336645, App1
40	32	88.9	62	17	US-10-425-115-259867	Sequence 259867, App1
41	32	88.9	62	17	US-10-425-115-283316	Sequence 283316, App1
42	32	88.9	66	17	US-10-425-115-220914	Sequence 220914, App1
43	32	88.9	68	15	US-10-424-589-158838	Sequence 158838, App1
44	32	88.9	71	17	US-10-425-115-332685	Sequence 332685, App1
45	32	88.9	73	15	US-10-424-589-285389	Sequence 285389, App1
46	32	88.9	76	15	US-10-424-589-239335	Sequence 239335, App1
47	32	88.9	76	17	US-10-425-115-349179	Sequence 349179, App1
48	32	88.9	78	17	US-10-425-115-325353	Sequence 325353, App1
49	32	88.9	81	17	US-10-425-115-320538	Sequence 320538, App1
50	32	88.9	82	17	US-10-425-115-193365	Sequence 193365, App1
51	32	88.9	85	17	US-10-425-115-295988	Sequence 295988, App1
52	32	88.9	86	16	US-10-437-963-134878	Sequence 134878, App1
53	32	88.9	89	9	US-09-739-907-83	Sequence 83, App1
54	32	88.9	89	11	US-09-938-671-182	Sequence 182, App1
55	32	88.9	91	15	US-10-289-762-726	Sequence 726, App1
56	32	88.9	100	17	US-10-425-115-209246	Sequence 209246, App1
57	32	88.9	102	17	US-10-425-115-303565	Sequence 303565, App1
58	32	88.9	104	17	US-10-425-115-334205	Sequence 334205, App1
59	32	88.9	106	15	US-10-424-589-284887	Sequence 284887, App1
60	32	88.9	114	15	US-10-424-589-146860	Sequence 146860, App1
61	32	88.9	116	15	US-10-424-589-172385	Sequence 172385, App1
62	32	88.9	116	17	US-10-425-115-293010	Sequence 293010, App1
63	32	88.9	118	9	US-09-739-907-182	Sequence 182, App1
64	32	88.9	118	11	US-09-938-671-182	Sequence 182, App1
65	32	88.9	120	16	US-10-767-701-57956	Sequence 57956, App1
66	32	88.9	124	17	US-10-425-115-201607	Sequence 201607, App1
67	32	88.9	126	15	US-10-424-589-235482	Sequence 235482, App1
68	32	88.9	132	17	US-10-425-115-185119	Sequence 185119, App1
69	32	88.9	132	17	US-10-425-115-340087	Sequence 340087, App1
70	32	88.9	142	17	US-10-425-115-313867	Sequence 313867, App1
71	32	88.9	143	15	US-10-424-589-225287	Sequence 225287, App1
72	32	88.9	144	17	US-10-425-115-280549	Sequence 280549, App1
73	32	88.9	148	17	US-10-425-115-299166	Sequence 299166, App1
74	32	88.9	152	15	US-10-424-589-251472	Sequence 251472, App1
75	32	88.9	155	17	US-10-425-115-253866	Sequence 253866, App1
76	32	88.9	157	16	US-10-437-963-119046	Sequence 119046, App1
77	32	88.9	164	16	US-10-437-963-119035	Sequence 119035, App1
78	32	88.9	166	16	US-10-437-963-102623	Sequence 102623, App1
79	32	88.9	166	16	US-10-437-963-174066	Sequence 174066, App1
80	32	88.9	174	16	US-10-437-963-119068	Sequence 119068, App1
81	32	88.9	186	16	US-10-767-701-322924	Sequence 322924, App1
82	32	88.9	194	14	US-10-104-047-2315	Sequence 2315, App1
83	32	88.9	198	9	US-09-995-553A-3	Sequence 5, App1
84	32	88.9	207	15	US-10-425-114-46073	Sequence 46073, App1
85	32	88.9	211	16	US-10-767-701-35984	Sequence 35984, App1
86	32	88.9	216	15	US-10-108-260A-4474	Sequence 4474, App1

```

87 32 88.9 229 14 US-10-156-761-12171 Sequence 12171, A
88 32 88.9 247 15 US-10-424-599-272621 Sequence 272621,
89 32 88.9 248 14 US-10-121-988-142 Sequence 142, App
90 32 88.9 248 14 US-10-200-562-142 Sequence 142, App
91 32 88.9 248 14 US-10-237-551-142 Sequence 142, App
92 32 88.9 249 17 US-10-739-930-7995 Sequence 7995, App
93 32 88.9 253 15 US-10-108-260A-3556 Sequence 3596, App
94 32 88.9 266 16 US-10-104-047-3114 Sequence 3114, App
95 32 88.9 269 16 US-10-437-963-169997 Sequence 169997,
96 32 88.9 275 15 US-10-329-148A-7 Sequence 7, Appli
97 32 88.9 280 17 US-10-493-462-31 Sequence 31, Appli
98 32 88.9 294 15 US-10-425-114-48469 Sequence 48469, A
99 32 88.9 296 17 US-10-763-362-6 Sequence 6, Appli
100 32 88.9 298 16 US-10-437-963-140406 Sequence 140406,
101 32 88.9 320 17 US-10-425-115-358651 Sequence 358651,
102 32 88.9 323 16 US-10-437-963-185078 Sequence 185078,
103 32 88.9 332 15 US-10-425-114-66862 Sequence 66862, A
104 32 88.9 338 16 US-10-437-963-186313 Sequence 186313,
105 32 88.9 351 14 US-10-104-047-2926 Sequence 2926, App
106 32 88.9 352 15 US-10-156-761-9067 Sequence 9067, App
107 32 88.9 362 15 US-10-425-114-62852 Sequence 62852, A
108 32 88.9 374 10 US-09-847-208-68 Sequence 68, Appli
109 32 88.9 393 17 US-10-739-930-10911 Sequence 10911, A
110 32 88.9 407 14 US-10-157-031-281 Sequence 281, App
111 32 88.9 424 13 US-10-029-180-104 Sequence 104, App
112 32 88.9 424 13 US-10-149-310-302 Sequence 302, App
113 32 88.9 430 16 US-10-767-701-45374 Sequence 45374, A
114 32 88.9 436 17 US-10-425-115-330241 Sequence 330241,
115 32 88.9 446 15 US-10-276-774-1370 Sequence 1370, App
116 32 88.9 462 15 US-10-282-122A-76568 Sequence 76568, A
117 32 88.9 531 14 US-10-369-493-2589 Sequence 2589, App
118 32 88.9 535 14 US-10-369-493-22630 Sequence 22630, A
119 32 88.9 581 14 US-10-094-466-40 Sequence 40, Appli
120 32 88.9 687 14 US-10-369-493-5922 Sequence 5922, App
121 32 88.9 699 14 US-10-121-988-143 Sequence 143, App
122 32 88.9 699 14 US-10-200-562-143 Sequence 143, App
123 32 88.9 699 14 US-10-237-551-143 Sequence 143, App
124 32 88.9 699 14 US-10-237-551-254 Sequence 254, App
125 32 88.9 719 14 US-10-156-761-9077 Sequence 9077, App
126 32 88.9 735 13 US-10-068-674-4 Sequence 4, Appli
127 32 88.9 773 16 US-10-437-963-121782 Sequence 121782,
128 32 88.9 799 16 US-10-437-963-145233 Sequence 145233,
129 32 88.9 804 14 US-10-236-392-58 Sequence 58, Appli
130 32 88.9 804 14 US-10-369-493-10624 Sequence 10624, A
131 32 88.9 878 15 US-10-369-493-197-1 Sequence 1, Appli
132 32 88.9 969 15 US-10-094-886-140 Sequence 140, App
133 32 88.9 1036 9 US-09-995-593A-6 Sequence 6, Appli
134 32 88.9 1067 17 US-10-650-650-18 Sequence 18, Appli
135 32 88.9 1187 9 US-09-995-593A-7 Sequence 7, Appli
136 32 88.9 1193 17 US-10-877-563-8 Sequence 8, Appli
137 32 88.9 1208 14 US-10-213-329-1 Sequence 1, Appli
138 32 88.9 1218 9 US-09-855-722-7 Sequence 7, Appli
139 32 88.9 1218 9 US-09-995-593A-11 Sequence 11, Appli
140 32 88.9 1218 9 US-09-944-849-3 Sequence 3, Appli
141 32 88.9 1218 14 US-10-219-248-7 Sequence 7, Appli
142 32 88.9 1218 14 US-10-219-247-7 Sequence 7, Appli
143 32 88.9 1218 14 US-10-281-478-7 Sequence 7, Appli
144 32 88.9 1218 15 US-10-236-392-56 Sequence 56, Appli
145 32 88.9 1218 16 US-10-648-593-155 Sequence 155, App
146 32 88.9 1218 17 US-10-763-362-43 Sequence 43, Appli
147 32 88.9 1218 17 US-10-877-563-4 Sequence 4, Appli
148 32 88.9 1218 17 US-10-650-650-1 Sequence 1, Appli
149 32 88.9 1404 9 US-09-944-849-8 Sequence 8, Appli
150 32 88.9 4840 14 US-10-156-761-10435 Sequence 10435, A

```

## ALIGNMENTS

RESULT 1  
US-10-719-385-26  
Sequence 26, Application US/10719385  
Publication No. US20040209284A1

```

; GENERAL INFORMATION:
; APPLICANT: O'Toole et al..
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-26

```

```

Query Match 100.0%; Score 36; DB 17; Length 6;
Best Local Similarity 100.0%; Pred. No. 1.5e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Query 1 AGGPGCV 6
Db 1 AGGPGCV 6

```

```

RESULT 2
US-10-425-115-359921
Sequence 359921, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53222)B
CURRENT FILING DATE: 2003-04-28
CURRENT APPLICATION NUMBER: US/10/425,115
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 359921
LENGTH: 68
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRF4577_9141C.1.pep
US-10-425-115-359921

```

```

Query Match 100.0%; Score 36; DB 17; Length 68;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Query 1 AGGPGCV 6
Db 38 AGGPGCV 43

```

```

RESULT 3
US-10-437-963-157034
Sequence 157034, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

```

```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 157034
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: clone ID: PAT_MKT4530_56644C.1.Dep
US-10-437-963-157034

Query Match
Best Local Similarity 100.0%; Score 36; DB 16; Length 339;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AGPCV 6
Db 6 AGPCV 11

RESULT 4
US-10-408-765A-817
; Sequence 817, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary W.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 817
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-817

Query Match
Best Local Similarity 100.0%; Score 36; DB 16; Length 407;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AGPCV 6
Db 280 AGPCV 285

RESULT 5
US-10-177-293-47
; Sequence 47, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Little, James
; APPLICANT: Glavic, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavairpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Meyer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
```

```

; APPLICANT: Baat Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Puzstai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Ayesgul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 858
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-47

Query Match
Best Local Similarity 100.0%; Score 36; DB 14; Length 858;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AGPCV 6
Db 731 AGPCV 736

RESULT 6
US-10-719-385-21
; Sequence 21, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1745
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-21

Query Match
Best Local Similarity 100.0%; Score 36; DB 17; Length 1745;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AGPCV 6
Db 1 AGPCV 6

RESULT 7
US-10-370-715B-544
```

```
; Sequence 544, Application US/10370715B
; Publication No. US20040258678A1
; GENERAL INFORMATION:
;   Patin Docket Preview
;   APPLICANT: BODARY, SARAH C.
;   APPLICANT: CLARK, HILARY
;   APPLICANT: BRISDELL, HUNTE
;   APPLICANT: JACKMAN, JANET
;   APPLICANT: SCHOENFELD, JILL R.
;   APPLICANT: WILLIAMS, P. MICKY
;   APPLICANT: WOOD, WILLIAM I.
;   APPLICANT: WU, THOMAS D.
; TITLE OF INVENTION: Compositions and Methods for the Treatment of Immune
; FILE REFERENCE: P1948P1-US
; CURRENT APPLICATION NUMBER: US/10/370,715B
; CURRENT FILING DATE: 2003-02-21
; NUMBER OF SEQ ID NOS: 742
; SEQ ID NO 544
; LENGTH: 1745
; TYPE: PRT
; ORGANISM: Homo sapien
; US-10-370-715B-544

Query Match      100.0%; Score 36; DB 17; Length 1745;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGPCV 6
       |||||
Db      1 AGPCV 6

RESULT 8
US-10-719-385-22
; Sequence 22, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
;   APPLICANT: O'TOOLE et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 3534
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-719-385-22

Query Match      100.0%; Score 36; DB 17; Length 3534;
Best Local Similarity 100.0%; Pred. No. 2.6e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGPCV 6
       |||||
Db      19 AGPCV 24

RESULT 9
US-10-424-599-230302
; Sequence 230302, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa Thomas J
;   APPLICANT: KOVALIC David K
;   APPLICANT: Zhou Yihua
;   APPLICANT: CAO Yongwei
; US-10-424-599-230302
```

```
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 230302
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_49987C.1.pep
; US-10-424-599-230302

Query Match      97.2%; Score 35; DB 15; Length 206;
Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGPCV 6
       |||||
Db      159 AGPCV 164

RESULT 10
US-10-424-599-149825
; Sequence 149825, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa Thomas J
;   APPLICANT: KOVALIC David K
;   APPLICANT: Zhou Yihua
;   APPLICANT: CAO Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 149825
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(378)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_106313C.1.pep
; US-10-424-599-149825

Query Match      97.2%; Score 35; DB 15; Length 378;
Best Local Similarity 83.3%; Pred. No. 5.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGPCV 6
       |||||
Db      131 AGPCV 136

RESULT 11
US-10-424-599-279664
; Sequence 279664, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa Thomas J
;   APPLICANT: KOVALIC David K
;   APPLICANT: Zhou Yihua
;   APPLICANT: CAO Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
```

NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 279664  
LENGTH: 409  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(409)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MKT3847\_94559C.1 pep  
US-10-424-599-279664

Query Match 97.2% Score 35; DB 15; Length 409;  
Best Local Similarity 83.3% Pred. No. 1.3e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 161 AGPCV 166

RESULT 12  
US-09-771-161A-255  
Sequence 255, Application US/09771161A  
Patent No. US20020110811A1  
GENERAL INFORMATION:  
APPLICANT: LEVINE, et al.  
TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES  
FILE REFERENCE: 802620-2005.1  
CURRENT FILING DATE: 2001-01-26  
PRIOR FILING DATE: 2000-04-12  
PRIOR APPLICATION NUMBER: 09/724,676  
PRIOR FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: 136776  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: 135619  
PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 273  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 255  
LENGTH: 1036  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-771-161A-255

Query Match 97.2% Score 35; DB 9; Length 1036;  
Best Local Similarity 83.3% Pred. No. 1.3e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 837 AGPCV 842

RESULT 13  
US-09-771-161A-256  
Sequence 256, Application US/09771161A  
Patent No. US20020110811A1  
GENERAL INFORMATION:  
APPLICANT: LEVINE, et al.  
TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES  
FILE REFERENCE: 802620-2005.1  
CURRENT FILING DATE: 2001-01-26  
PRIOR FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: 09/724,676  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: 136776  
PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 273  
SOFTWARE: Patentin version 3.0

SEQ ID NO 256  
LENGTH: 1036  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-771-161A-256

Query Match 97.2% Score 35; DB 9; Length 1036;  
Best Local Similarity 83.3% Pred. No. 1.3e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 837 AGPCV 842

RESULT 14  
US-10-318-410-2  
Sequence 2, Application US/10318410  
Publication No. US20030175763A1  
GENERAL INFORMATION:  
APPLICANT: Yan Y. DEGENHARDT  
APPLICANT: Scott POWERS  
TITLE OF INVENTION: IDENTIFICATION OF AN AMPLIFIED GENE AND TARGET FOR DRUG  
FILE REFERENCE: 38002-0040  
CURRENT FILING DATE: 2002-12-13  
PRIOR APPLICATION NUMBER: US/10/318,410  
PRIOR FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 2  
LENGTH: 1036  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-318-410-2

Query Match 97.2% Score 35; DB 14; Length 1036;  
Best Local Similarity 83.3% Pred. No. 1.3e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 837 AGPCV 842

RESULT 15  
US-10-737-450-72  
Sequence 72, Application US/10737450  
Publication No. US20040235071A1  
GENERAL INFORMATION:  
APPLICANT: Millennium Pharmaceuticals, Inc  
APPLICANT: Lightcap, Eric S.  
APPLICANT: Ecsedy, Jeffrey A.  
APPLICANT: Hunter, John Joseph  
APPLICANT: Macbeth, Kyle J  
APPLICANT: Tighe Nestor, Michelle  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING  
TITLE OF INVENTION: CANCER USING 15986, 2188, 20743, 9148, 9151, 9791, 44252,  
TITLE OF INVENTION: 14184, 42461, 8204, 7970, 2552, 21657, 26492, 2411, 15088,  
TITLE OF INVENTION: 1905, 28899, 63380, 33935, 10480, 12686, 25501, 17694,  
TITLE OF INVENTION: 15701, 53062, 49908, 21612, 38949, 6216, 46863, 9235, 2201,  
TITLE OF INVENTION: 6985, 9883, 12238, 10057, 21617, 39228, 49928, 54476, 62113,  
TITLE OF INVENTION: 64316, 12264, 32362, 58198, 2887, 3205, 8557, 9600, 9693,  
TITLE OF INVENTION: 44867, 53058, 55556, 57658, 2208, 10252, 10302, 14218,  
TITLE OF INVENTION: 33877, 10317, 10485, 25964, 14815, 1363, 1397, 14827, 21708,  
FILE REFERENCE: MP102-207P1R0X0N1M  
CURRENT FILING DATE: 2003-12-16  
PRIOR APPLICATION NUMBER: US/10/737,450  
PRIOR FILING DATE: 2002-12-20  
PRIOR APPLICATION NUMBER: US 60/436,443

```

; PRIOR FILING DATE: 2002-12-23
; PRIOR APPLICATION NUMBER: US 60/438,498
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/444,370
; PRIOR FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: US 60/446,031
; PRIOR FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: US 60/453,635
; PRIOR FILING DATE: 2003-03-11
; PRIOR APPLICATION NUMBER: US 60/457,199
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: US 60/462,458
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US 60/466,732
; PRIOR FILING DATE: 2003-04-30
; PRIOR APPLICATION NUMBER: US 60/469,184
; PRIOR FILING DATE: 2003-05-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1092
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-737-450-72

Query Match      97.2%; Score 35; DB 17; Length 1092;
Best Local Similarity 83.3%; Pred. No. 1.4e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGPCV 6
        |||||:
Db      893 AGPCL 898

RESULT 16
US-10-437-963-201950
; Sequence 201950, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 201950
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(82)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_97276C.1.pep
US-10-437-963-201950

Query Match      91.7%; Score 33; DB 16; Length 82;
Best Local Similarity 83.3%; Pred. No. 3.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGPCV 6
        |||||:
Db      15 AGPCL 20

US-10-437-963-153250
; Sequence 153250, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 153250
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53220C.1.pep
US-10-437-963-153250

Query Match      91.7%; Score 33; DB 16; Length 93;
Best Local Similarity 83.3%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGPCV 6
        |||||:
Db      80 AGPCL 85

RESULT 18
US-10-437-963-121278
; Sequence 121278, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 121278
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_24319C.1.pep
US-10-437-963-121278

Query Match      91.7%; Score 33; DB 16; Length 95;
Best Local Similarity 83.3%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGPCV 6
        |||||:
Db      70 AGPCL 75
```

```
RESULT 19
US-10-767-701-51998
; Sequence 51998, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 51998
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB5121-002-A1-PFI-E10.pep
US-10-767-701-51998

Query Match          91.7%; Score 33; DB 16; Length 98;
Best Local Similarity 83.3%; Pred. No. 3.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

CY      1 AGPCV 6
      |||||:
Db      32 AGPCL 37

RESULT 20
US-10-437-963-170381
; Sequence 170381, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 170381
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_68713C.1.pcp
US-10-437-963-170381

Query Match          91.7%; Score 33; DB 16; Length 149;
Best Local Similarity 83.3%; Pred. No. 5.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

CY      1 AGPCV 6
      |||||:
Db      80 AGPCL 85

RESULT 21
US-10-437-963-129834
; Sequence 129834, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
```

```
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 129834
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(152)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_32052C.1.pcp
US-10-437-963-129834

Query Match          91.7%; Score 33; DB 16; Length 152;
Best Local Similarity 83.3%; Pred. No. 5.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

CY      1 AGPCV 6
      |||||:
Db      84 AGPCL 89

RESULT 22
US-10-425-115-235622
; Sequence 235622, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 235622
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_146471C.1.pcp
US-10-425-115-235622

Query Match          91.7%; Score 33; DB 17; Length 338;
Best Local Similarity 83.3%; Pred. No. 1.1e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

CY      1 AGPCV 6
      |||||:
Db      9 AGPCL 14

RESULT 23
US-10-767-701-40379
; Sequence 40379, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
```



APPLICANT: Zhou, Yihua  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 40379  
; LENGTH: 351  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: SORBI-28WAY03-C58371\_1.pep  
US-10-767-701-40379

Query Match 91.7%; Score 33; DB 16; Length 351;  
Best Local Similarity 83.3%; Pred. No. 1.1e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 8 AGPCV 13

RESULT 24  
US-10-767-701-41600  
; Sequence 41600, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 41600  
; LENGTH: 355  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: SORBI-28WAY03-C50197\_1.pep  
US-10-767-701-41600

Query Match 91.7%; Score 33; DB 16; Length 355;  
Best Local Similarity 83.3%; Pred. No. 1.1e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 9 AGPCV 14

RESULT 25  
US-10-425-114-52879  
; Sequence 52879, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 52879  
; LENGTH: 371  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3060-051-G8\_FIL.pep  
US-10-425-114-52879

Query Match 91.7%; Score 33; DB 15; Length 371;  
Best Local Similarity 83.3%; Pred. No. 1.2e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 42 AGPCV 47

RESULT 26  
US-10-437-963-180870  
; Sequence 180870, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 180870  
; LENGTH: 782  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_7819C.1.pep  
US-10-437-963-180870

Query Match 91.7%; Score 33; DB 16; Length 782;  
Best Local Similarity 83.3%; Pred. No. 2.2e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 6 AGPCV 11

RESULT 27  
US-10-437-963-156470  
; Sequence 156470, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 156470

LENGTH: 931  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(931)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_56134C.1.pdp  
US-10-437-963-156470

Query Match 91.7%; Score 33; DB 16; Length 931;  
Best Local Similarity 83.3%; Pred. No. 2.6e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPCV 6  
DB 307 AGPCV 312

RESULT 28  
US-10-437-963-194642  
Sequence 194642, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 194642  
LENGTH: 1824  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(1824)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_90668C.1.pdp  
US-10-437-963-194642

Query Match 91.7%; Score 33; DB 16; Length 1824;  
Best Local Similarity 83.3%; Pred. No. 4.6e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPCV 6  
DB 1775 SGPCV 1780

RESULT 29  
US-10-354-240-34  
Sequence 34, Application US/10354240  
Publication No. US20030185847A1  
GENERAL INFORMATION:  
APPLICANT: Sone, Toshio  
APPLICANT: Kume, Akinori  
APPLICANT: Dairiki, Kazuo  
APPLICANT: Iwama, Akiyo  
APPLICANT: Kino, Kohsuke  
TITLE OF INVENTION: Peptide-Based Immunotherapeutic Agent for Treating Allergic Diseases  
FILE REFERENCE: SPO-103D1  
CURRENT APPLICATION NUMBER: US/10/354,240

CURRENT FILING DATE: 2003-01-29  
PRIOR APPLICATION NUMBER: PCT/JP97/00740  
PRIOR FILING DATE: 1997-03-10  
PRIOR APPLICATION NUMBER: US 09/142,524  
PRIOR FILING DATE: 1998-09-09  
NUMBER OF SEQ ID NOS: 174  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 34  
LENGTH: 15  
TYPE: PRT  
ORGANISM: Cryptomeria japonica  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (1)..(15)  
OTHER INFORMATION: Cryj1 peptide, Figure 1, Row 120  
US-10-354-240-34

Query Match 88.9%; Score 32; DB 14; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
DB 9 GGPCV 13

RESULT 30  
US-10-354-240-35  
Sequence 35, Application US/10354240  
Publication No. US20030185847A1  
GENERAL INFORMATION:  
APPLICANT: Sone, Toshio  
APPLICANT: Kume, Akinori  
APPLICANT: Dairiki, Kazuo  
APPLICANT: Iwama, Akiyo  
APPLICANT: Kino, Kohsuke  
TITLE OF INVENTION: Peptide-Based Immunotherapeutic Agent for Treating Allergic Diseases  
FILE REFERENCE: SPO-103D1  
CURRENT APPLICATION NUMBER: US/10/354,240  
CURRENT FILING DATE: 2003-01-29  
PRIOR APPLICATION NUMBER: PCT/JP97/00740  
PRIOR FILING DATE: 1997-03-10  
PRIOR APPLICATION NUMBER: US 09/142,524  
PRIOR FILING DATE: 1998-09-09  
NUMBER OF SEQ ID NOS: 174  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 35  
LENGTH: 15  
TYPE: PRT  
ORGANISM: Cryptomeria japonica  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (1)..(15)  
OTHER INFORMATION: Cryj1 peptide, Figure 1, Row 21  
US-10-354-240-35

Query Match 88.9%; Score 32; DB 14; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
DB 4 GGPCV 8

RESULT 31  
US-10-425-115-341764  
Sequence 341764, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 341764  
LENGTH: 38  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_74857C.1.pep  
US-10-425-115-341764

Query Match  
Best Local Similarity 88.9%; Score 32; DB 17; Length 38;  
Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGPCV 6  
Db 8 GGPCV 12

RESULT 32  
US-10-437-963-186690  
Sequence 186690, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Bouharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 186690  
LENGTH: 43  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_83466C.1.pep  
US-10-437-963-186690

Query Match  
Best Local Similarity 88.9%; Score 32; DB 16; Length 43;  
Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPC 5  
Db 25 AGPC 29

RESULT 33  
US-10-425-115-249558  
Sequence 249558, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115

CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 249558  
LENGTH: 46  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_159181C.1.pep  
US-10-425-115-249558

Query Match  
Best Local Similarity 88.9%; Score 32; DB 17; Length 46;  
Pred. No. 2.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGPCV 6  
Db 39 GGPCV 43

RESULT 34  
US-10-029-386-31343  
Sequence 31343, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Hanzel, David R.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
FILE REFERENCE: AEOMICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 31343  
LENGTH: 48  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC012384.16  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4  
OTHER INFORMATION: SWISSPROT HIT: P21851, EVALUATE 1.70e+00  
US-10-029-386-31343

Query Match  
Best Local Similarity 88.9%; Score 32; DB 14; Length 48;  
Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGPCV 6  
Db 39 GGPCV 43

RESULT 35  
US-10-425-115-286513  
Sequence 286513, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 286513  
LENGTH: 48  
TYPE: PRT

ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_24398C.1.pep  
US-10-425-115-286513

Query Match 88.9%; Score 32; DB 17; Length 48;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPCV 6  
DB 20 SGPCV 25

RESULT 36  
US-10-424-599-241100  
Sequence 241100, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 241100  
LENGTH: 52  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_59741C.1.pep  
US-10-424-599-241100

Query Match 88.9%; Score 32; DB 15; Length 52;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPCV 5  
DB 38 AGPCV 42

RESULT 37  
US-10-424-599-181045  
Sequence 181045, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 181045  
LENGTH: 54  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1) (54)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_13449C.1.pep  
US-10-424-599-181045

Query Match 88.9%; Score 32; DB 15; Length 54;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGPCV 6  
DB 38 GGPCV 42

RESULT 38  
US-10-424-599-152102  
Sequence 152102, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 152102  
LENGTH: 57  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1) (57)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_108371C.1.pep  
US-10-424-599-152102

Query Match 88.9%; Score 32; DB 15; Length 57;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGPCV 6  
DB 20 GGPCV 24

RESULT 39  
US-10-425-115-336645  
Sequence 336645, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 336645  
LENGTH: 57  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_70149C.1.pep  
US-10-425-115-336645

Query Match 88.9%; Score 32; DB 17; Length 57;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPCV 5

Db 48 AGPC 52

RESULT 40

US-10-425-115-259867  
 ; Sequence 259867, Application US/10425115  
 ; Publication No. US20040214272A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants  
 ; FILE REFERENCE: 38-21(53222)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,115  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 369326  
 ; SEQ ID NO 259867  
 ; LENGTH: 62  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: MRT4577\_168617C.1.pep  
 US-10-425-115-259867

Query Match 88.9%; Score 32; DB 17; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGPC 5  
 Db 36 AGPC 40

Search completed: December 30, 2004, 15:31:16  
 Job time : 4.73678 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 30, 2004, 15:14:08 ; Search time 215.263 Seconds

(without alignments)  
2929.449 Million cell updates/sec

Title: US-10-719-385-2

Perfect score: 9007  
Sequence: 1 MIRSKITSVLSFCRSSREL.....PEGEPLIQVQAVRMQR 1753Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1599051 seqs, 359727711 residues

Total number of hits satisfying chosen parameters: 1599051

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 150 summaries

## Database :

Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	9007	100.0	1753	17	US-10-719-385-2	Sequence 2, Appli
2	9005	100.0	1753	17	US-10-719-385-4	Sequence 4, Appli
3	9005	100.0	1753	17	US-10-719-385-8	Sequence 8, Appli
4	9004	100.0	1753	17	US-10-719-385-3	Sequence 3, Appli
5	9004	100.0	1753	17	US-10-719-385-11	Sequence 11, Appli
6	9004	100.0	1753	17	US-10-719-385-13	Sequence 13, Appli
7	9004	100.0	1753	17	US-10-719-385-14	Sequence 14, Appli
8	9003	100.0	1753	17	US-10-719-385-15	Sequence 15, Appli
9	9002	99.9	1753	17	US-10-719-385-6	Sequence 6, Appli
10	9002	99.9	1753	17	US-10-719-385-10	Sequence 10, Appli
11	9001	99.9	1753	17	US-10-719-385-16	Sequence 16, Appli
12	8999	99.9	1753	17	US-10-719-385-7	Sequence 7, Appli
13	8999	99.9	1753	17	US-10-719-385-12	Sequence 12, Appli

14	8998	99.9	1753	17	US-10-719-385-18	Sequence 18, Appli
15	8997	99.9	1753	17	US-10-719-385-9	Sequence 9, Appli
16	8993	99.8	1753	17	US-10-719-385-17	Sequence 17, Appli
17	8991	99.8	1753	17	US-10-719-385-19	Sequence 19, Appli
18	8985.5	99.8	1752	17	US-10-719-385-5	Sequence 5, Appli
19	8939	99.2	1745	17	US-10-719-385-21	Sequence 21, Appli
20	8939	99.2	1745	17	US-10-370-715B-544	Sequence 544, Appli
21	8927	99.1	1745	17	US-10-719-385-22	Sequence 22, Appli
22	5270	58.5	1111	17	US-10-719-385-23	Sequence 23, Appli
23	4093	45.4	853	17	US-10-719-385-24	Sequence 24, Appli
24	2420	26.9	525	17	US-10-719-385-25	Sequence 25, Appli
25	369.5	4.1	2026	16	US-10-437-963-135530	Sequence 135530, Appli
26	305	3.4	63	14	US-10-106-698-5917	Sequence 5917, Ap
27	245	2.7	1745	16	US-10-437-963-173637	Sequence 173637, Ap
28	176.5	2.0	2905	16	US-10-437-963-154118	Sequence 154118, Ap
29	175.5	1.9	1676	14	US-10-128-714-8246	Sequence 8246, Ap
30	169.5	1.9	2621	16	US-10-437-963-122168	Sequence 122168, Ap
31	166	1.8	2122	16	US-10-437-963-189782	Sequence 189782, Ap
32	165	1.8	1545	14	US-10-128-714-3246	Sequence 3246, Ap
33	160.5	1.8	2462	16	US-10-437-963-114113	Sequence 114113, Ap
34	157	1.7	3859	16	US-10-408-765A-354	Sequence 354, Ap
35	156	1.7	2627	15	US-10-424-599-256710	Sequence 256710, Ap
36	155	1.7	1557	14	US-10-369-493-2224	Sequence 2224, Ap
37	155	1.7	2834	15	US-10-424-599-256711	Sequence 256711, Ap
38	153	1.7	3830	17	US-10-723-860-2568	Sequence 2568, Ap
39	152.5	1.7	2159	16	US-10-437-963-108860	Sequence 108860, Ap
40	152.5	1.7	2412	16	US-10-408-765A-214	Sequence 214, Ap
41	151	1.7	3225	16	US-10-408-765A-254	Sequence 254, Ap
42	151	1.7	4464	14	US-10-369-493-5019	Sequence 5019, Ap
43	148.5	1.6	2811	17	US-10-810-352-39	Sequence 39, Appli
44	148	1.6	2593	16	US-10-437-963-114115	Sequence 114115, Ap
45	147.5	1.6	2209	9	US-09-902-941-1903	Sequence 1903, Ap
46	147.5	1.6	2209	9	US-09-849-626-1903	Sequence 1903, Ap
47	147.5	1.6	2209	14	US-10-017-754-1903	Sequence 1903, Ap
48	147.5	1.6	2209	14	US-10-113-872-1903	Sequence 1903, Ap
49	147.5	1.6	2209	14	US-10-283-017-1903	Sequence 883, App
50	147.5	1.6	2209	16	US-10-408-765A-863	Sequence 1675, Ap
51	147.5	1.6	2209	17	US-10-723-860-1675	Sequence 249, App
52	147.5	1.6	2859	13	US-10-087-192-249	Sequence 2, Appli
53	147	1.6	3907	14	US-10-171-111-2	Sequence 648, Ap
54	147	1.6	3925	14	US-10-171-111-6	Sequence 6048, Ap
55	145.5	1.6	2848	14	US-10-369-493-6048	Sequence 10532, A
56	145	1.6	2311	17	US-10-739-930-10532	Sequence 3, Appli
57	144.5	1.6	2549	10	US-09-950-634-3	Sequence 3, Appli
58	144.5	1.6	2549	16	US-10-688-016-3	Sequence 16815, Ap
59	144.5	1.6	2405	9	US-10-437-963-106615	Sequence 4789, Ap
60	143.5	1.6	876	15	US-09-738-626-4789	Sequence 86, Appli
61	143	1.6	1783	15	US-10-188-832-86	Sequence 2, Appli
62	143	1.6	2549	16	US-10-701-490-2	Sequence 4, Appli
63	143	1.6	3899	14	US-10-171-111-4	Sequence 8, Appli
64	143	1.6	3917	14	US-10-171-111-8	Sequence 18286, Ap
65	142.5	1.6	973	16	US-10-437-963-178286	Sequence 76, Appli
66	142.5	1.6	2545	15	US-10-092-900A-76	Sequence 189923, Ap
67	142	1.6	2646	16	US-10-437-963-189923	Sequence 203003, Ap
68	140.5	1.6	1821	16	US-10-437-963-203003	Sequence 14, Appli
69	140.5	1.6	2852	17	US-10-481-471-14	Sequence 117806, Ap
70	140	1.6	1867	16	US-10-437-963-137806	Sequence 11, Appli
71	140	1.6	3878	14	US-10-080-608A-11	Sequence 100, App
72	140	1.6	3911	16	US-10-408-765A-1839	Sequence 1839, App
73	140	1.6	3911	17	US-10-473-127-571	Sequence 415, App
74	139.5	1.5	4096	15	US-10-363-616-415	Sequence 569, App
75	139.5	1.5	4097	16	US-10-408-765A-1585	Sequence 574, App
76	139.5	1.5	4097	17	US-10-473-127-569	Sequence 215, App
77	139.5	1.5	4097	17	US-10-363-616-416	Sequence 161362, Ap
78	139.5	1.5	4128	15	US-10-473-127-574	Sequence 4, Appli
79	139.5	1.5	4128	17	US-10-473-127-574	Sequence 572, App
80	138.5	1.5	1651	17	US-10-128-558-215	Sequence 106122, Ap
81	138.5	1.5	2132	15	US-10-424-599-161362	Sequence 568, App
82	138.5	1.5	2835	9	US-09-885-535-4	Sequence 573, App
83	138.5	1.5	4095	16	US-10-473-127-572	
84	138	1.5	1479	17	US-10-437-963-106122	
85	136.5	1.5	4128	17	US-10-473-127-568	
86	136.5	1.5	4128	17	US-10-473-127-573	

87 135.5 1.5 2335 14 US-10-369-493-2561 Sequence 2561, Ap  
88 135 1.5 1230 14 US-10-205-219-44 Sequence 44, Ap  
89 135 1.5 1417 16 US-10-205-219-97 Sequence 97, Ap  
90 134.5 1.5 1417 16 US-10-408-765A-671 Sequence 671, Ap  
91 134.5 1.5 2222 14 US-10-369-493-3923 Sequence 3923, Ap  
92 134 1.5 1905 15 US-10-259-194A-86 Sequence 86, Ap  
93 134 1.5 2470 14 US-10-369-493-22191 Sequence 22191, A  
94 134 1.5 5183 14 US-10-107-521-1 Sequence 1, Ap  
95 133.5 1.5 789 16 US-10-437-963-119795 Sequence 119795,  
96 133.5 1.5 1088 13 US-10-001-867-127 Sequence 127, Ap  
97 133.5 1.5 1088 14 US-10-074-475-255 Sequence 255, Ap  
98 133 1.5 1494 14 US-10-195-144-17 Sequence 17, Ap  
99 133 1.5 1494 14 US-10-345-072-17 Sequence 11, Ap  
100 132.5 1.5 1837 14 US-10-369-493-22734 Sequence 22734, A  
101 131.5 1.5 2055 13 US-10-017-216-4 Sequence 4, Ap  
102 131.5 1.5 2151 15 US-10-424-599-161366 Sequence 161366,  
103 131 1.5 2221 15 US-10-424-599-161365 Sequence 161365,  
104 131 1.5 2677 16 US-10-367-094-124 Sequence 124, Ap  
105 131 1.5 4128 14 US-10-205-194-1 Sequence 1, Ap  
106 130 1.4 1819 16 US-10-437-963-111842 Sequence 111842,  
107 130 1.4 2111 16 US-10-437-963-125597 Sequence 125597,  
108 129.5 1.4 1818 14 US-10-032-585-7616 Sequence 7616, Ap  
109 129.5 1.4 1826 15 US-10-289-762-113 Sequence 113, Ap  
110 129 1.4 1966 14 US-10-369-493-6049 Sequence 6049, Ap  
111 128.5 1.4 1826 15 US-10-282-122A-54869 Sequence 54869, A  
112 128.5 1.4 3092 9 US-09-801-368-172 Sequence 172, Ap  
113 128.5 1.4 3092 14 US-10-369-493-1470 Sequence 1470, Ap  
114 128 1.4 1166 17 US-10-370-715B-562 Sequence 562, Ap  
115 128 1.4 2458 14 US-10-281-172-257 Sequence 257, Ap  
116 128 1.4 2458 15 US-10-221-278-257 Sequence 257, Ap  
117 128 1.4 1960 15 US-10-210-172-36 Sequence 36, Ap  
118 127.5 1.4 2517 15 US-10-313-203-27 Sequence 27, Ap  
119 127.5 1.4 3433 16 US-10-408-765A-731 Sequence 731, Ap  
120 127.5 1.4 3433 16 US-10-408-765A-732 Sequence 732, Ap  
121 127.5 1.4 3829 16 US-10-408-765A-849 Sequence 849, Ap  
122 127.5 1.4 3829 17 US-10-723-860-2137 Sequence 2137, Ap  
123 127.5 1.4 3863 15 US-10-197-824-7 Sequence 7, Ap  
124 127.5 1.4 5171 16 US-10-408-765A-2687 Sequence 2687, Ap  
125 127.5 1.4 1542 16 US-10-408-765A-269 Sequence 269, Ap  
126 126.5 1.4 808 14 US-10-408-765A-442 Sequence 442, Ap  
127 126.5 1.4 2671 16 US-10-104-047-2324 Sequence 2324, Ap  
128 126 1.4 925 16 US-10-369-493-28281 Sequence 28281, Ap  
129 126 1.4 4025 16 US-10-437-963-193926 Sequence 193926,  
130 126 1.4 5373 16 US-10-408-765A-741 Sequence 741, Ap  
131 126 1.4 920 16 US-10-437-963-136710 Sequence 136710,  
132 125.5 1.4 1377 14 US-10-205-342-25 Sequence 25, Ap  
133 125.5 1.4 1464 17 US-10-331-053-83 Sequence 83, Ap  
134 125.5 1.4 2541 14 US-10-177-293-470 Sequence 470, Ap  
135 125.5 1.4 1933 17 US-10-723-860-1971 Sequence 1971, Ap  
136 125 1.4 916 16 US-10-473-574-22 Sequence 22, Ap  
137 124.5 1.4 1475 14 US-10-007-047-2 Sequence 2, Ap  
138 124.5 1.4 821 14 US-10-369-493-2791 Sequence 2791, Ap  
139 124 1.4 1209 15 US-10-282-122A-67611 Sequence 67611, A  
140 124 1.4 2457 16 US-10-437-963-115622 Sequence 115622,  
141 124 1.4 4313 15 US-10-287-226-312 Sequence 312, Ap  
142 124 1.4 1193 16 US-10-437-963-185505 Sequence 185505,  
143 123.5 1.4 1283 14 US-10-007-047-4 Sequence 4, Ap  
144 123.5 1.4 1451 10 US-09-988-027-1 Sequence 1, Ap  
145 123.5 1.4 1451 14 US-10-165-099-1 Sequence 1, Ap  
146 123.5 1.4 1451 14 US-09-998-027-4 Sequence 4, Ap  
147 123.5 1.4 1471 10 US-10-165-099-4 Sequence 4, Ap  
148 123.5 1.4 1471 11 US-10-964-956-11 Sequence 11, Ap  
149 123.5 1.4 2053 14 US-09-964-956-11 Sequence 11, Ap  
150 123.5 1.4 2053 15 US-10-262-511-2 Sequence 2, Ap

## ALIGNMENTS

GENERAL INFORMATION:  
; APPLICANT: O'Toole et al.  
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
; FILE REFERENCE: 22058-582  
; CURRENT APPLICATION NUMBER: US/10/719,385  
; CURRENT FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: PCT/US03/37339  
; PRIOR FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: 60/428,094  
; PRIOR FILING DATE: 2002-11-21  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1753  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-719-385-2  
Query Match 100.0%; Score 9007; DB 17; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MIRSKITSVLSFCRSSRELMTLLGRSALRELQTEAEINKMRRLLBGLSYKPPSPS 60  
Db 1 MIRSKITSVLSFCRSSRELMTLLGRSALRELQTEAEINKMRRLLBGLSYKPPSPS 60  
QY 61 SAEVYKANKVAPLKLGLRISKFLGLDEQSVOLLQCYLOEDYKGTBRSVKTVLQDER 120  
Db 61 SAEVYKANKVAPLKLGLRISKFLGLDEQSVOLLQCYLOEDYKGTBRSVKTVLQDER 120  
QY 121 QSQALILIKADYYEERTCILRCVHLITLYFQDERHPYREYVADCVKLEKELVSKYRQ 180  
Db 121 QSQALILIKADYYEERTCILRCVHLITLYFQDERHPYREYVADCVKLEKELVSKYRQ 180  
QY 181 PEEIYKTEAPTWETHGNLMTEROVSRWFVQCLREQSMLEIIFLYAYFEMASDILVLT 240  
Db 181 PEEIYKTEAPTWETHGNLMTEROVSRWFVQCLREQSMLEIIFLYAYFEMASDILVLT 240  
QY 241 KMFEEQSGSRQNRHNVDETMDFVDRIGYFALILVEQMDIESLHKALDDBREIHOF 300  
Db 241 KMFEEQSGSRQNRHNVDETMDFVDRIGYFALILVEQMDIESLHKALDDBREIHOF 300  
QY 301 AODGLICQDDCMILTFGDIPEHAPVYLAWALRHTLNPEETSVMKIGTATQLVNFQ 360  
Db 301 AODGLICQDDCMILTFGDIPEHAPVYLAWALRHTLNPEETSVMKIGTATQLVNFQ 360  
QY 361 YLTRLQSLASGNDCTTSTACMCVGLSFVLTSLHTLGNQODIIDRACEVLADPSL 420  
Db 361 YLTRLQSLASGNDCTTSTACMCVGLSFVLTSLHTLGNQODIIDRACEVLADPSL 420  
QY 421 PELFWGTEPISGLIILDSVCGMFPHLSPLQLALVSGKSTAKKVSFLKMSFVNE 480  
Db 421 PELFWGTEPISGLIILDSVCGMFPHLSPLQLALVSGKSTAKKVSFLKMSFVNE 480  
QY 481 LYHKHPDIVSHBDGTLMRQRTKLLYPLGGQTNLRIPOGTQVQWMLDDBAYLVREYSY 540  
Db 481 LYHKHPDIVSHBDGTLMRQRTKLLYPLGGQTNLRIPOGTQVQWMLDDBAYLVREYSY 540  
QY 541 SSMTLFCELEMLLHVSTADVQHOCQRYKPIIDLVHKYISTDLSTADCLPITSRYML 600  
Db 541 SSMTLFCELEMLLHVSTADVQHOCQRYKPIIDLVHKYISTDLSTADCLPITSRYML 600  
QY 601 LQRLTVISPPVIVASCVNCLTVLAARPAKWTOLRHGFLPFAHPVSSISOMISAE 660  
Db 601 LQRLTVISPPVIVASCVNCLTVLAARPAKWTOLRHGFLPFAHPVSSISOMISAE 660  
QY 661 GNAAGGYNLMLNSQPOGEYGVITAFRLITTLVKGQSGTSQGLVPCVMEVLKEMLP 720  
Db 661 GNAAGGYNLMLNSQPOGEYGVITAFRLITTLVKGQSGTSQGLVPCVMEVLKEMLP 720  
QY 721 SYHKRWYNSHGVAEQIGCLILIELIHALIHLCHETDLSHSTBLOFLCICSLAYTEAGOT 780  
Db 721 SYHKRWYNSHGVAEQIGCLILIELIHALIHLCHETDLSHSTBLOFLCICSLAYTEAGOT 780



```

QY 781 VININGIGVDITDMMAOPRSDGAEQGGGLIKTKYLAFTSYNNYIRLKPENNVSP 840
Db 781 VININGIGVDITDMMAOPRSDGAEQGGGLIKTKYLAFTSYNNYIRLKPENNVSP 840
QY 841 LEQALSOQHAGNNMLIAVLAKIYKAKHPALPRLAIOQLKRLATVAPMSVYACLGNDAA 900
Db 841 LEQALSOQHAGNNMLIAVLAKIYKAKHPALPRLAIOQLKRLATVAPMSVYACLGNDAA 900
QY 901 IRDAFLTRLOSKIEDMRIRKWMLEFLVAVETOPGLIELFLNLEVKDSDGSKESFLGMW 960
Db 901 IRDAFLTRLOSKIEDMRIRKWMLEFLVAVETOPGLIELFLNLEVKDSDGSKESFLGMW 960
QY 961 SCILAVLILDSQQODRYWCPEPLHRAAIAFLHLMQORRDSAMLVTRKPFENLTS 1020
Db 961 SCILAVLILDSQQODRYWCPEPLHRAAIAFLHLMQORRDSAMLVTRKPFENLTS 1020
QY 1021 LFGTSPSESTSEPSILETCALIMKICLEIYVVVKSIDOSLKDOLKFKSIEKFPAYMS 1080
Db 1021 LFGTSPSESTSEPSILETCALIMKICLEIYVVVKSIDOSLKDOLKFKSIEKFPAYMS 1080
QY 1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSAMRMLIIATTADIMHLDTSVVRQLFLDV 1140
Db 1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSAMRMLIIATTADIMHLDTSVVRQLFLDV 1140
QY 1141 LDGTALLVPASVNCIALGSKCTLLIILRQWRELGSDVEILGPILEGLQADQ 1200
Db 1141 LDGTALLVPASVNCIALGSKCTLLIILRQWRELGSDVEILGPILEGLQADQ 1200
QY 1201 QLMETKAKVSAFTTVQMKEMKYSIDIPQYSQVLVANCETLOEVIALLFOTRISLALG 1260
Db 1201 QLMETKAKVSAFTTVQMKEMKYSIDIPQYSQVLVANCETLOEVIALLFOTRISLALG 1260
QY 1261 SATEDKDSMETDDCSRRHRDQDVCYGLHLAELCEVEDGDSWLOVTRRLPILPTL 1320
Db 1261 SATEDKDSMETDDCSRRHRDQDVCYGLHLAELCEVEDGDSWLOVTRRLPILPTL 1320
QY 1321 LTTLEVSIRMKONHFTATLHLLTLARQOGAVAVAGAGITOSICPLSVYOLSTNG 1380
Db 1321 LTTLEVSIRMKONHFTATLHLLTLARQOGAVAVAGAGITOSICPLSVYOLSTNG 1380
QY 1381 TAQTSASRKSIDASMPGCVYRLSMSLEBOLKTLRYNPLPALDPFGVHQBERTLOCLNA 1440
Db 1381 TAQTSASRKSIDASMPGCVYRLSMSLEBOLKTLRYNPLPALDPFGVHQBERTLOCLNA 1440
QY 1441 VRTVQSLACLEBADHTVGFIIQLSNFMKEWHFLPQLMRDIQVNLGYLCOACTSLHSRK 1500
Db 1441 VRTVQSLACLEBADHTVGFIIQLSNFMKEWHFLPQLMRDIQVNLGYLCOACTSLHSRK 1500
QY 1501 MLOHVLQNKNGDGLPSAVAOQVORPPSAASAAPSSSKOPADTEASEOQALHTVQYGLK 1560
Db 1501 MLOHVLQNKNGDGLPSAVAOQVORPPSAASAAPSSSKOPADTEASEOQALHTVQYGLK 1560
QY 1561 ILSTKLALRHFTPVCOILIDQSLDLAEYNFLPALSTPTPTFDESVAPSFSTLLATVNV 1620
Db 1561 ILSTKLALRHFTPVCOILIDQSLDLAEYNFLPALSTPTPTFDESVAPSFSTLLATVNV 1620
QY 1621 ALNMLGELDKKKEPLTOAVGLSTOAEGRITLKSLLMFTMENC FYLLISQAMRYLDPANH 1680
Db 1621 ALNMLGELDKKKEPLTOAVGLSTOAEGRITLKSLLMFTMENC FYLLISQAMRYLDPANH 1680
QY 1681 PRDKRMQOEJLSSELSTLSSLSRYFRGAPSSPATVLPSPQKSTLSKASPSPOEPL 1740
Db 1681 PRDKRMQOEJLSSELSTLSSLSRYFRGAPSSPATVLPSPQKSTLSKASPSPOEPL 1740
QY 1741 IQLVOAFVRHMQR 1753
Db 1741 IQLVOAFVRHMQR 1753

```

RESULT 2  
US-10-719-385-4  
; Sequence 4, Application US/10719385

```

; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-4

Query Match 100.0%; Score 9005; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFCRSSRELWTILLGRSALRELQIEALNKHWRLLLEGSYKPPSPS 60
Db 1 MIRSKITSVLSFCRSSRELWTILLGRSALRELQIEALNKHWRLLLEGSYKPPSPS 60
QY 61 SAEKVKANKDVASPLKEIGLRISKFLGLDEFOQVOLLCCYIOEDYGRGTRSDVKTVALDER 120
Db 61 SAEKVKANKDVASPLKEIGLRISKFLGLDEFOQVOLLCCYIOEDYGRGTRSDVKTVALDER 120
QY 121 OSQALILKADYVEERFCILRCVHLITVRODEHPRVAVAPCVDLREKIVSKYRQO 180
Db 121 OSQALILKADYVEERFCILRCVHLITVRODEHPRVAVAPCVDLREKIVSKYRQO 180
QY 181 FEELYKTEAPMETHNGMLNTEROVSRMFVOCLEQESMLLEIIFLYAYFENAPSDLVLT 240
Db 181 FEELYKTEAPMETHNGMLNTEROVSRMFVOCLEQESMLLEIIFLYAYFENAPSDLVLT 240
QY 241 KMFKEQGFSGQTRNHLVDETMDFVDRIGYFSALILVEGMDIESLHCALDDRRREHOF 300
Db 241 KMFKEQGFSGQTRNHLVDETMDFVDRIGYFSALILVEGMDIESLHCALDDRRREHOF 300
QY 301 AODGLICODMCMLETFGDIPIHHAIVLAAMLLRHTLNPETSSVVRKIGGTAIOLNVQ 360
Db 301 AODGLICODMCMLETFGDIPIHHAIVLAAMLLRHTLNPETSSVVRKIGGTAIOLNVQ 360
QY 361 YLTRLQSLAGSNDCTSTACMCYGLSFTLSLEHTLGNQODIIDTACEVLADPSL 420
Db 361 YLTRLQSLAGSNDCTSTACMCYGLSFTLSLEHTLGNQODIIDTACEVLADPSL 420
QY 421 PELFWGTEPTSGLIILDSVCGMPHLLSPILQLRALVSGKSTAKVYSFLDKMSFYNE 480
Db 421 PELFWGTEPTSGLIILDSVCGMPHLLSPILQLRALVSGKSTAKVYSFLDKMSFYNE 480
QY 481 LYKHKPHVISHEDGTLNRQTPKLLYPLGGQTNIRIQGVVGVOMLDDRAYLVKWEYSY 540
Db 481 LYKHKPHVISHEDGTLNRQTPKLLYPLGGQTNIRIQGVVGVOMLDDRAYLVKWEYSY 540
QY 541 SSWTLFCEIEMLHVSTADVIQHCQVKKPIDLVHXYITDLSIADCLPITSRIYML 600
Db 541 SSWTLFCEIEMLHVSTADVIQHCQVKKPIDLVHXYITDLSIADCLPITSRIYML 600
QY 601 LORLTTVISPVDVIASCNCCLTVLAANPAKWTDLRHGTLPFVAHPVSSLSQMSIAE 660
Db 601 LORLTTVISPVDVIASCNCCLTVLAANPAKWTDLRHGTLPFVAHPVSSLSQMSIAE 660
QY 661 GNNAGGYNLLMNSRQGEYGVTTAFLRLITTLVKGQSGTSQSGVLPVCMFVLKEMLP 720
Db 661 GNNAGGYNLLMNSRQGEYGVTTAFLRLITTLVKGQSGTSQSGVLPVCMFVLKEMLP 720
QY 721 SYHKRVYNSHGREQIGCLIELHAIINLGHETDLHSHTPSLOFLCISLAYTEAQOT 780

```

```

Db 721 SYHKRYSNGVREQIGCILLILHAIINLCHETDLSHSHHTPSLOFLCISLATEAGOT 780
Qy 761 VININGIGVDTIDWMAAOPRSDGAEQGGQGLIKTVKAFSVNNVIRLKPENVSP 840
Db 761 VININGIGVDTIDWMAAOPRSDGAEQGGQGLIKTVKAFSVNNVIRLKPENVSP 840
Qy 841 LEQALSOGAGNNLIVLAKYIYKHDPALPRALIOILLKRLATVAPSVYACLGNDAAA 900
Db 841 LEQALSOGAGNNLIVLAKYIYKHDPALPRALIOILLKRLATVAPSVYACLGNDAAA 900
Qy 901 IRDAFLTLRQSKIEEMRIKMLLEFLTVAVETQPGILFELFNLVYKQSDSKESLGMW 960
Db 901 IRDAFLTLRQSKIEEMRIKMLLEFLTVAVETQPGILFELFNLVYKQSDSKESLGMW 960
Qy 961 SCIAHVLILDSQODRWCPPLHRAAIAFLHAIAMODRBSAMVLRKPFEMENLSP 1020
Db 961 SCIAHVLILDSQODRWCPPLHRAAIAFLHAIAMODRBSAMVLRKPFEMENLSP 1020
Qy 1021 LFGTSPSESTSEPSILETCALIMKICLEIYVVYKGLDQSLKDTLKFSIEKRFAYMS 1080
Db 1021 LFGTSPSESTSEPSILETCALIMKICLEIYVVYKGLDQSLKDTLKFSIEKRFAYMS 1080
Qy 1081 GYVSLAHVAETESSCTSLSEYOMLVSAARMMLIITTHADIMHLDTSVVRQLFDV 1140
Db 1081 GYVSLAHVAETESSCTSLSEYOMLVSAARMMLIITTHADIMHLDTSVVRQLFDV 1140
Qy 1141 LDGTALLVPAVAVNCLRLGSMKCTLLILRLQMKRELSVDEILGPTTEILEGVLQDQ 1200
Db 1141 LDGTALLVPAVAVNCLRLGSMKCTLLILRLQMKRELSVDEILGPTTEILEGVLQDQ 1200
Qy 1201 QLMKTKKVPFAFTVLQMKEMKVS DIPQYSQVLVNCETLQOEIVIALPDQTHSLAG 1260
Db 1201 QLMKTKKVPFAFTVLQMKEMKVS DIPQYSQVLVNCETLQOEIVIALPDQTHSLAG 1260
Qy 1261 SATEDKSMETDSCRSRHRDORGVCLGIHAKELCEVEDDEGDSMTQRRRLPIPLTL 1320
Db 1261 SATEDKSMETDSCRSRHRDORGVCLGIHAKELCEVEDDEGDSMTQRRRLPIPLTL 1320
Qy 1321 LTTLEVSJRMKONLHFEATLHLTLTLARQOGATAVAGAGITQSLCPLSVYQLSTNG 1380
Db 1321 LTTLEVSJRMKONLHFEATLHLTLTLARQOGATAVAGAGITQSLCPLSVYQLSTNG 1380
Qy 1381 TAQTPSARSKSLDAPSWRGVYRLSMLSEOLIKTRNPLRBEALDFGVCHOERTLOCLNA 1440
Db 1381 TAQTPSARSKSLDAPSWRGVYRLSMLSEOLIKTRNPLRBEALDFGVCHOERTLOCLNA 1440
Qy 1441 VRTVQSLACLBEADHTVGFILQLSNFMKEWHFHLPOLMRDIOVNLGYLCOACTSLHSRK 1500
Db 1441 VRTVQSLACLBEADHTVGFILQLSNFMKEWHFHLPOLMRDIOVNLGYLCOACTSLHSRK 1500
Qy 1501 MLQHYLQNKNGDGLPSAVAQEVORPSPASASAPSSKQPAADTEASEQALHTVQYGLK 1560
Db 1501 MLQHYLQNKNGDGLPSAVAQEVORPSPASASAPSSKQPAADTEASEQALHTVQYGLK 1560
Qy 1561 ILSTLIALRHFPTDVQCILLDOSIDLAEVNFALSTPTTPPSEVAPSGITLAVNV 1620
Db 1561 ILSTLIALRHFPTDVQCILLDOSIDLAEVNFALSTPTTPPSEVAPSGITLAVNV 1620
Qy 1621 ALANLGEIDKKEPELTOAVGSLTOAEGRTLSLIMFMENCFFULLSQAMRYLRDPAVH 1680
Db 1621 ALANLGEIDKKEPELTOAVGSLTOAEGRTLSLIMFMENCFFULLSQAMRYLRDPAVH 1680
Qy 1681 PRDQRMKQKELSSSLSTLSSLSRYFRGAPSSPATVLPSPQKSTSLSKASPESEPL 1740
Db 1681 PRDQRMKQKELSSSLSTLSSLSRYFRGAPSSPATVLPSPQKSTSLSKASPESEPL 1740
Qy 1741 IQVQAFVRHMOR 1753
Db 1741 IQVQAFVRHMOR 1753

```

RESULT 3  
US-10-719-385-8

```

; Sequence 8, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-8

Query Match 100.0%; Score 9005; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRSKITSVLSPCRSSREIMTLILGRSALRELSQIEAEINKMRRLLEGSLYKPPSPS 60
Db 1 MIRSKITSVLSPCRSSREIMTLILGRSALRELSQIEAEINKMRRLLEGSLYKPPSPS 60
Qy 61 SAEYKANKQVAPLKELELRISKFLGLDEBSVOLIOCYLOEDYRGTDSVXTVLODER 120
Db 61 SAEYKANKQVAPLKELELRISKFLGLDEBSVOLIOCYLOEDYRGTDSVXTVLODER 120
Qy 121 QSQALIKIADYYEERTCLRCVLHLTYFOOERHPYRYEVDQVCKLEKELYSKXROQ 180
Db 121 QSQALIKIADYYEERTCLRCVLHLTYFOOERHPYRYEVDQVCKLEKELYSKXROQ 180
Qy 181 FEELYKTEADPTWETHGNLMTEROVSFVQCLREQSMLEIIFLYYAFFEMAPSDLLVLT 240
Db 181 FEELYKTEADPTWETHGNLMTEROVSFVQCLREQSMLEIIFLYYAFFEMAPSDLLVLT 240
Qy 241 KMFKEQFGSRQNRHLVDETMDPVDRIQYFSALILVEEMDIESLHKCALDRREIHOF 300
Db 241 KMFKEQFGSRQNRHLVDETMDPVDRIQYFSALILVEEMDIESLHKCALDRREIHOF 300
Qy 301 AODGLICQDMDCLMTFEGDIPHHAPVILAWALRHTLNPEETSVVRKIGTAIQLVNFQ 360
Db 301 AODGLICQDMDCLMTFEGDIPHHAPVILAWALRHTLNPEETSVVRKIGTAIQLVNFQ 360
Qy 361 YLTRLQSLASGNDCTTSTACMCVYGLSFVLSLELHTLGNQODIIDTACEVLADPSL 420
Db 361 YLTRLQSLASGNDCTTSTACMCVYGLSFVLSLELHTLGNQODIIDTACEVLADPSL 420
Qy 421 PELFMGTEPISGILLDSVCGMFPHLSPLOLALVSGSTAKKVSFLDKMSFYNE 480
Db 421 PELFMGTEPISGILLDSVCGMFPHLSPLOLALVSGSTAKKVSFLDKMSFYNE 480
Qy 481 LYHKKHDVISHEDGTLMRRQPKLXPLGGQNNLAIPOGTQVQWLDDBAYLVREHYSY 540
Db 481 LYHKKHDVISHEDGTLMRRQPKLXPLGGQNNLAIPOGTQVQWLDDBAYLVREHYSY 540
Qy 541 SSWTLFTCELEMIHLHVSADVDVQHOCQRYKPIIDLVHXYISTDLSTADCLLPTISRIMYL 600
Db 541 SSWTLFTCELEMIHLHVSADVDVQHOCQRYKPIIDLVHXYISTDLSTADCLLPTISRIMYL 600
Qy 601 LQRLTVISPPVDVIVASCVNCLTVLAARPAKWTDLRHTGFLPFAHVPSSLSQMSAE 660
Db 601 LQRLTVISPPVDVIVASCVNCLTVLAARPAKWTDLRHTGFLPFAHVPSSLSQMSAE 660
Qy 661 GNNAGYGNLNNSEPOGEYGVITAFRLITTLVVGQIGSTOSQGLVPCVMEVLEMLP 720
Db 661 GNNAGYGNLNNSEPOGEYGVITAFRLITTLVVGQIGSTOSQGLVPCVMEVLEMLP 720
Qy 721 SYHKRYSNGVREQIGCILLILHAIINLCHETDLSHSHHTPSLOFLCISLATEAGOT 780

```

```

Db 721 SYHKRYSHGREGIGCLIELHAIILNLGHETDHSHTPSIQFLICISLAETAGQT 780
Qy 781 VINIMGICVDITIDWMAAQPSRSDAGEGGOGOLIKYTKALPSTYNNVIRLKPSPNVSP 840
Db 781 VINIMGICVDITIDWMAAQPSRSDAGEGGOGOLIKYTKALPSTYNNVIRLKPSPNVSP 840
Qy 841 LEQALSOHGAGNNLIATLAKYIYKHDPALPRLATOLLKRLATAPWVSACGNDAAA 900
Db 841 LEQALSOHGAGNNLIATLAKYIYKHDPALPRLATOLLKRLATAPWVSACGNDAAA 900
Qy 901 IRDAFLRLQSKIEDMRKIMLEFLJVAVEVTOPLIELFNLLEVKGSDSKESLGMW 960
Db 901 IRDAFLRLQSKIEDMRKIMLEFLJVAVEVTOPLIELFNLLEVKGSDSKESLGMW 960
Qy 961 SCLHVLVELINSOQDRYWCPLLRHRAIAFLHLMORRSMVLVATKRFMENTLSP 1020
Db 961 SCLHVLVELINSOQDRYWCPLLRHRAIAFLHLMORRSMVLVATKRFMENTLSP 1020
Qy 1021 LFGTSPSETSEPSILETCALIMKICLIEIYVYVKSGLDOSLKDYLKFKFSIEKRFAYMS 1080
Db 1021 LFGTSPSETSEPSILETCALIMKICLIEIYVYVKSGLDOSLKDYLKFKFSIEKRFAYMS 1080
Qy 1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSAMRMIIATTHADIMHLDTSVVRQLFLDV 1140
Db 1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSAMRMIIATTHADIMHLDTSVVRQLFLDV 1140
Qy 1141 LDGKALILVPAVAVNCCLAGSMKCTLLILRLKMRKELSGVDEILGPIITELLEGVLQDQ 1200
Db 1141 LDGKALILVPAVAVNCCLAGSMKCTLLILRLKMRKELSGVDEILGPIITELLEGVLQDQ 1200
Qy 1201 QLMKTKYAVSAFTVLQMKMKVSDIPOYSOLVANCETLQOEVIAPOTRHSALG 1260
Db 1201 QLMKTKYAVSAFTVLQMKMKVSDIPOYSOLVANCETLQOEVIAPOTRHSALG 1260
Qy 1261 SATEDKOSMETDSCSRSHRDQDQCVLGLHLAKELCEVEDGDGSMVLQVTRRLPIEPL 1320
Db 1261 SATEDKOSMETDSCSRSHRDQDQCVLGLHLAKELCEVEDGDGSMVLQVTRRLPIEPL 1320
Qy 1321 LTTLEVSIRKMONLHFTBATLHLITLARTOOGATVAGAGITOSTICPLISVYQOLSTNG 1380
Db 1321 LTTLEVSIRKMONLHFTBATLHLITLARTOOGATVAGAGITOSTICPLISVYQOLSTNG 1380
Qy 1381 TAQPSASRSKSLDAPSWGVYRLSWSLMEOLLKTLRYFLPEALDPFGVGHERTLOCINA 1440
Db 1381 TAQPSASRSKSLDAPSWGVYRLSWSLMEOLLKTLRYFLPEALDPFGVGHERTLOCINA 1440
Qy 1441 VRTVQSLACLEADHTVGFIIQLSNFMKEMHFLPOLNRDIQVNLGYLCOACTSLHSRK 1500
Db 1441 VRTVQSLACLEADHTVGFIIQLSNFMKEMHFLPOLNRDIQVNLGYLCOACTSLHSRK 1500
Qy 1501 MLQHYLQNKNGDGLPSAVARQVPPSASAPSSSKOPPAADTEASECOALHTVOYGLK 1560
Db 1501 MLQHYLQNKNGDGLPSAVARQVPPSASAPSSSKOPPAADTEASECOALHTVOYGLK 1560
Qy 1561 ILSKTLALRHFTPPVCOIILDOSJDLABYNFLFSLSTTTPPDBEVAPSFGTLATVNV 1620
Db 1561 ILSKTLALRHFTPPVCOIILDOSJDLABYNFLFSLSTTTPPDBEVAPSFGTLATVNV 1620
Qy 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRITLSLMTMENC FYLLISQAMRYLADPAVH 1680
Db 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRITLSLMTMENC FYLLISQAMRYLADPAVH 1680
Qy 1681 PRDKQRMQOELSESELSTLSLSRYFRGAPSSPATVLPSPQKSTLSLQASPSQSPPL 1740
Db 1681 PRDKQRMQOELSESELSTLSLSRYFRGAPSSPATVLPSPQKSTLSLQASPSQSPPL 1740
Qy 1741 IQLVQAFVRHMOR 1753
Db 1741 IQLVQAFVRHMOR 1753

```

RESULT 4

```

US-10-719-385-3
; Sequence 3, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1753
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-719-385-3

```

```

Query Match 100.0%; Score 9004; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 MIRSKITSVLSFCRSSRELTTILGRSALRELQIEAEIKHMRRLLEGSTYKPPSPS 60
Db 1 MIRSKITSVLSFCRSSRELTTILGRSALRELQIEAEIKHMRRLLEGSTYKPPSPS 60
Qy 61 SAEKYANKADVAPLKEIGLRISKFLGDEBQSVOLLQCYLOEDYRGTRDSVKTVLQDER 120
Db 61 SAEKYANKADVAPLKEIGLRISKFLGDEBQSVOLLQCYLOEDYRGTRDSVKTVLQDER 120
Qy 121 QSOALLIKIADYYEERTCIRCTYHLITTYQDERHRYRYEYADCVLKEKLSYKTRQO 180
Db 121 QSOALLIKIADYYEERTCIRCTYHLITTYQDERHRYRYEYADCVLKEKLSYKTRQO 180
Qy 181 FEELYKTEAPTWETHNGMLNTEROYSRMFVOCLEQOSMLLEIIFLYYAYFEWAPSDLVLT 240
Db 181 FEELYKTEAPTWETHNGMLNTEROYSRMFVOCLEQOSMLLEIIFLYYAYFEWAPSDLVLT 240
Qy 241 KMFEEGGRSROTNRHLVDETMDFVDRIGYFSALILVEGNDIBSLHKCALDDRREHLQF 300
Db 241 KMFEEGGRSROTNRHLVDETMDFVDRIGYFSALILVEGNDIBSLHKCALDDRREHLQF 300
Qy 301 AODGLICODMCLMTGEDI PHHAPVLLAWALLHNTLNPESTSSVVRKIGGTALQLVNFQ 360
Db 301 AODGLICODMCLMTGEDI PHHAPVLLAWALLHNTLNPESTSSVVRKIGGTALQLVNFQ 360
Qy 361 YLTRLGSLASGGNDCTTSTACMCVYGLSFLVLSLELHTLGNODIIDTACEVLADPSL 420
Db 361 YLTRLGSLASGGNDCTTSTACMCVYGLSFLVLSLELHTLGNODIIDTACEVLADPSL 420
Qy 421 PELFWGTEPTSGGLIILDSVCGMFPHLSPILLQRLALVSGSKTAKVYSFLDMRSPFNE 480
Db 421 PELFWGTEPTSGGLIILDSVCGMFPHLSPILLQRLALVSGSKTAKVYSFLDMRSPFNE 480
Qy 481 LYKXKPHDVISHBEGTLMRRQTPKLLYPLGGQTMIRIPQGVGQVMDDRAYLYRWREYSY 540
Db 481 LYKXKPHDVISHBEGTLMRRQTPKLLYPLGGQTMIRIPQGVGQVMDDRAYLYRWREYSY 540
Qy 541 SSMWLFCEIEMLAHVSTADVIOHCORVKPIIDLVHKVISTDLSIADCLPITSRIYML 600
Db 541 SSMWLFCEIEMLAHVSTADVIOHCORVKPIIDLVHKVISTDLSIADCLPITSRIYML 600
Qy 601 LQRLITVISPVDVYIASCVCNLTVLAARNPAKWTDLRHGTFLEFVAHPVSSLQMSIAE 660
Db 601 LQRLITVISPVDVYIASCVCNLTVLAARNPAKWTDLRHGTFLEFVAHPVSSLQMSIAE 660
Qy 661 GMANGGYGNLIMNSBOGEGYGTIATFLRLITTYVKQGLSTQSGVLPVCMFYLKEMLP 720
Db 661 GMANGGYGNLIMNSBOGEGYGTIATFLRLITTYVKQGLSTQSGVLPVCMFYLKEMLP 720

```

QY 721 SYHKRNYSHGVRQIGCLIELIELIAINLCHETDLSHSHPSLOFLCISLATEAGOT 780  
 DB 721 SYHKRNYSHGVRQIGCLIELIELIAINLCHETDLSHSHPSLOFLCISLATEAGOT 780  
 QY 781 VININGIGVDTIDMMAAOPRSDGABGOGGOLIKTVKLAFSVTNNVIRLKPSPNVSP 840  
 DB 781 VININGIGVDTIDMMAAOPRSDGABGOGGOLIKTVKLAFSVTNNVIRLKPSPNVSP 840  
 QY 841 VININGIGVDTIDMMAAOPRSDGABGOGGOLIKTVKLAFSVTNNVIRLKPSPNVSP 840  
 DB 841 VININGIGVDTIDMMAAOPRSDGABGOGGOLIKTVKLAFSVTNNVIRLKPSPNVSP 840  
 QY 841 VININGIGVDTIDMMAAOPRSDGABGOGGOLIKTVKLAFSVTNNVIRLKPSPNVSP 840  
 DB 841 VININGIGVDTIDMMAAOPRSDGABGOGGOLIKTVKLAFSVTNNVIRLKPSPNVSP 840  
 QY 901 IRDAFLTLQSKIEDMRIKVMILEFLVAVETOPGLIELFLNLEVKDSDSKESLGMW 960  
 DB 901 IRDAFLTLQSKIEDMRIKVMILEFLVAVETOPGLIELFLNLEVKDSDSKESLGMW 960  
 QY 901 IRDAFLTLQSKIEDMRIKVMILEFLVAVETOPGLIELFLNLEVKDSDSKESLGMW 960  
 DB 901 IRDAFLTLQSKIEDMRIKVMILEFLVAVETOPGLIELFLNLEVKDSDSKESLGMW 960  
 QY 961 SCINAVLELIDSOODRWCPPLHRAIAFLHAIEMORRBSANLVLTKEKMENTLSP 1020  
 DB 961 SCINAVLELIDSOODRWCPPLHRAIAFLHAIEMORRBSANLVLTKEKMENTLSP 1020  
 QY 961 SCINAVLELIDSOODRWCPPLHRAIAFLHAIEMORRBSANLVLTKEKMENTLSP 1020  
 DB 961 SCINAVLELIDSOODRWCPPLHRAIAFLHAIEMORRBSANLVLTKEKMENTLSP 1020  
 QY 1021 LFGTLPSPSETSEPSILETCALIMKICLIEIYVVKSLDOSLMDTLKKFSIEKRFAYWS 1080  
 DB 1021 LFGTLPSPSETSEPSILETCALIMKICLIEIYVVKSLDOSLMDTLKKFSIEKRFAYWS 1080  
 QY 1021 LFGTLPSPSETSEPSILETCALIMKICLIEIYVVKSLDOSLMDTLKKFSIEKRFAYWS 1080  
 DB 1021 LFGTLPSPSETSEPSILETCALIMKICLIEIYVVKSLDOSLMDTLKKFSIEKRFAYWS 1080  
 QY 1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSANRMLIIATTHADIMHLDTSVVRQLFLDV 1140  
 DB 1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSANRMLIIATTHADIMHLDTSVVRQLFLDV 1140  
 QY 1141 LDGTALLVPAVNVCLFLGSMKCTLLIILRQWREIGSDVDELIGPTLLEGLVQADQ 1200  
 DB 1141 LDGTALLVPAVNVCLFLGSMKCTLLIILRQWREIGSDVDELIGPTLLEGLVQADQ 1200  
 QY 1201 QLMETKAKVSAFITTVOEMKEMKSDIPQYSQVLAVNCETLOEBVIALFPQTHSLALG 1260  
 DB 1201 QLMETKAKVSAFITTVOEMKEMKSDIPQYSQVLAVNCETLOEBVIALFPQTHSLALG 1260  
 QY 1261 SATEDKSMETDDCSRRHRDQDVCVGLHAIKELCEVDEBQSDQVTRRLPIPTL 1320  
 DB 1261 SATEDKSMETDDCSRRHRDQDVCVGLHAIKELCEVDEBQSDQVTRRLPIPTL 1320  
 QY 1321 LTTLEVSILRMKONHFTBATLHLTLTLARTOGATAVAGAGITOSICLPILSVQSLTNG 1380  
 DB 1321 LTTLEVSILRMKONHFTBATLHLTLTLARTOGATAVAGAGITOSICLPILSVQSLTNG 1380  
 QY 1381 TAQTPSASRKSILDAFSPMGVYRLSMSLMEQLKTLRYNLFPEALDPVGHQERTLQCLNA 1440  
 DB 1381 TAQTPSASRKSILDAFSPMGVYRLSMSLMEQLKTLRYNLFPEALDPVGHQERTLQCLNA 1440  
 QY 1441 VRTVQSLACLBEADHTVGFIIQLSNFMKEMHFLPOLMRDIQVNLGYLCACTSLHSRK 1500  
 DB 1441 VRTVQSLACLBEADHTVGFIIQLSNFMKEMHFLPOLMRDIQVNLGYLCACTSLHSRK 1500  
 QY 1501 MLQHYLQWKNQDGLPSAFAORVORPPSAASAAPSSKOPADTASBEOALHTVQYGLK 1560  
 DB 1501 MLQHYLQWKNQDGLPSAFAORVORPPSAASAAPSSKOPADTASBEOALHTVQYGLK 1560  
 QY 1561 ILSTKTLAALRHFTPDVCOILLDOSLDAEYVFLFALSTPTPTPSEVAPSGTLLAVNV 1620  
 DB 1561 ILSTKTLAALRHFTPDVCOILLDOSLDAEYVFLFALSTPTPTPSEVAPSGTLLAVNV 1620  
 QY 1621 ALNMLGELDKKKEPILTOAVGISTOAEGRITKSLIMTMENCFLIILSOAMRYLADVAH 1680  
 DB 1621 ALNMLGELDKKKEPILTOAVGISTOAEGRITKSLIMTMENCFLIILSOAMRYLADVAH 1680  
 QY 1681 PRDQRMKQELSESLTSLSSRYFRGAPSPATGVLPSPOKSTSLSKASPESQEP 1740  
 DB 1681 PRDQRMKQELSESLTSLSSRYFRGAPSPATGVLPSPOKSTSLSKASPESQEP 1740  
 QY 1741 IOLVOAFVRHQR 1753  
 DB 1741 IOLVOAFVRHQR 1753

RESULT 5  
 US-10-719-385-11  
 ; Sequence 11, Application US/10719385  
 ; Publication No. US20040209284A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Toole et al.  
 ; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 ; FILE REFERENCE: 22058-582  
 ; CURRENT APPLICATION NUMBER: US/10/719,385  
 ; PRIOR FILING DATE: 2003-11-21  
 ; PRIOR APPLICATION NUMBER: 60/428,094  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 11  
 ; LENGTH: 1753  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-719-385-11

Query Match 100.0%; Score 9004; DB 17; Length 1753;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRKSTSVLSFRCRSREIMTLLGRSALREISOIEALNKMRRLBGLSYKPPSPS 60  
 DB 1 MIRKSTSVLSFRCRSREIMTLLGRSALREISOIEALNKMRRLBGLSYKPPSPS 60  
 QY 61 SAEKRVANKOVASPLKELGLRISKFLGLDEBQSVOLLOCYLQBDYGRTRDSVTVLQDER 120  
 DB 61 SAEKRVANKOVASPLKELGLRISKFLGLDEBQSVOLLOCYLQBDYGRTRDSVTVLQDER 120  
 QY 121 QSQALLIKTADYYEERTCICRCVHLILTYFQDERHPRYEVADCVDKLEKELVSKYRQ 180  
 DB 121 QSQALLIKTADYYEERTCICRCVHLILTYFQDERHPRYEVADCVDKLEKELVSKYRQ 180  
 QY 181 FEELYTEAPWTETHGNLWTEROVSRMFVQCLREQSMLEIIFYLYAAYFEMAPSDLLVLT 240  
 DB 181 FEELYTEAPWTETHGNLWTEROVSRMFVQCLREQSMLEIIFYLYAAYFEMAPSDLLVLT 240  
 QY 241 KMFKEQFGSRQTRNLVDETMDPFVDRIGYFSALLIVEGMDIESLHKCALDDBRELHOF 300  
 DB 241 KMFKEQFGSRQTRNLVDETMDPFVDRIGYFSALLIVEGMDIESLHKCALDDBRELHOF 300  
 QY 301 AODGLICQDDCMLTFGDI PHHAPVLLAMALLRHTLNEBETSVVRKIGTAIQLVNQ 360  
 DB 301 AODGLICQDDCMLTFGDI PHHAPVLLAMALLRHTLNEBETSVVRKIGTAIQLVNQ 360  
 QY 361 YLTRLIQSLASGNDCTSTACMCVYGLSFVLTSLDELHTLGNQODIIDTACEVLADPSL 420  
 DB 361 YLTRLIQSLASGNDCTSTACMCVYGLSFVLTSLDELHTLGNQODIIDTACEVLADPSL 420  
 QY 421 PELFWGTEPTSGIGIILDSVCGMFPHLSPLLQLLRALVSGSTAKVYSFLDKMSFYNE 480  
 DB 421 PELFWGTEPTSGIGIILDSVCGMFPHLSPLLQLLRALVSGSTAKVYSFLDKMSFYNE 480  
 QY 481 LYGKREHDVISHEDGTLMRORPKLYPLGQONTLIPQCTGQVWLDDBRAYLVREYSY 540  
 DB 481 LYGKREHDVISHEDGTLMRORPKLYPLGQONTLIPQCTGQVWLDDBRAYLVREYSY 540  
 QY 541 SSWTLFTCEIEMLHVSTADVLOHCORVPIIIDLVHAKVISTDLSTADCLLPTSRIVNL 600  
 DB 541 SSWTLFTCEIEMLHVSTADVLOHCORVPIIIDLVHAKVISTDLSTADCLLPTSRIVNL 600  
 QY 601 LQRLTVISPPVDVIAASCNCLTVLAARNPAKWTDLRHTGFLPFVAHPVSSLSQMSIAE 660  
 DB 601 LQRLTVISPPVDVIAASCNCLTVLAARNPAKWTDLRHTGFLPFVAHPVSSLSQMSIAE 660  
 QY 661 GNNAGGYNLNNSBOGEYGVTTAFLALITTVKVGOLSGTOSQGLVPCVAFVLKEMLP 720  
 DB 661 GNNAGGYNLNNSBOGEYGVTTAFLALITTVKVGOLSGTOSQGLVPCVAFVLKEMLP 720

QY 721 SYHKRVNHSRGVREIOGCLILELTHAINTLCHETDLSHSHTPSLOFLCISLAYTEAGOT 780  
DB 721 SYHKRVNHSRGVREIOGCLILELTHAINTLCHETDLSHSHTPSLOFLCISLAYTEAGOT 780  
QY 781 VINIMIGVDTIDWMAAQPBSDGAEQGGQOLIKIVYKAFSVYTNVIRLKPSPNVSP 840  
DB 781 VINIMIGVDTIDWMAAQPBSDGAEQGGQOLIKIVYKAFSVYTNVIRLKPSPNVSP 840  
QY 841 LEQALSQGHAGNNLIATLAKYIYKHDPALPRALIQLKSLATVAPMSVYACIGNDAAA 900  
DB 841 LEQALSQGHAGNNLIATLAKYIYKHDPALPRALIQLKSLATVAPMSVYACIGNDAAA 900  
QY 901 IRDAPFLRLOSKIEDMRKIMILERLTAVENTOPGIELFINLEVGQSDSKSPSLGMM 960  
DB 901 IRDAPFLRLOSKIEDMRKIMILERLTAVENTOPGIELFINLEVGQSDSKSPSLGMM 960  
QY 961 SCLHVAVELIDSOQODRYWCPEPLHRAAIAFLHALMODRDSAMLVATKREPMENLTSP 1020  
DB 961 SCLHVAVELIDSOQODRYWCPEPLHRAAIAFLHALMODRDSAMLVATKREPMENLTSP 1020  
QY 1021 LFGTILSPSESTSEPSILETCALIMKICLETIYVYVKGSLDQSLKDTLKFSIEKFAWMS 1080  
DB 1021 LFGTILSPSESTSEPSILETCALIMKICLETIYVYVKGSLDQSLKDTLKFSIEKFAWMS 1080  
QY 1081 GYVKSIAVHVAETGSSCTSLLEYOMLVSAWRMLIITTHADIMHLDTSVVRQLFDV 1140  
DB 1081 GYVKSIAVHVAETGSSCTSLLEYOMLVSAWRMLIITTHADIMHLDTSVVRQLFDV 1140  
QY 1141 LDGFKALLVPASVNCRLGSMKCTLLILRLQWRRELSVDEILGPTTEILEGVLQADQ 1200  
DB 1141 LDGFKALLVPASVNCRLGSMKCTLLILRLQWRRELSVDEILGPTTEILEGVLQADQ 1200  
QY 1201 QLMKTKAKVPSAFTTVLQMKMKVSDIPQYSQVLVANCETLQOEVIALLPQTRHSLAG 1260  
DB 1201 QLMKTKAKVPSAFTTVLQMKMKVSDIPQYSQVLVANCETLQOEVIALLPQTRHSLAG 1260  
QY 1261 SATEDKOMETDPCSRSHRQRODGVCLGLHAKELCEVDEBDGSMVQVTRRLPIPLTL 1320  
DB 1261 SATEDKOMETDPCSRSHRQRODGVCLGLHAKELCEVDEBDGSMVQVTRRLPIPLTL 1320  
QY 1321 LTTLEVSILRMKONLHTEATLHLTLTLARTOQATAVAGAGITQSIICPLLSVYQOLSTNG 1380  
DB 1321 LTTLEVSILRMKONLHTEATLHLTLTLARTOQATAVAGAGITQSIICPLLSVYQOLSTNG 1380  
QY 1381 TAQTPSASRKSUDAPSPGVYRLSWSLMEQLIKTLRYNPLPEALDFVGVOERTLQCLNA 1440  
DB 1381 TAQTPSASRKSUDAPSPGVYRLSWSLMEQLIKTLRYNPLPEALDFVGVOERTLQCLNA 1440  
QY 1441 VRTYOSLACEADHTVGFIIQLSNFMKEMHFLPOLMRDIOVNLGYLCOACTSLHSRK 1500  
DB 1441 VRTYOSLACEADHTVGFIIQLSNFMKEMHFLPOLMRDIOVNLGYLCOACTSLHSRK 1500  
QY 1501 MLQHYLQKNGDGLPSAAVQRPSPASAAAPSSSKOPADTEASEGOALHTVOYGLK 1560  
DB 1501 MLQHYLQKNGDGLPSAAVQRPSPASAAAPSSSKOPADTEASEGOALHTVOYGLK 1560  
QY 1561 ILSTKLALRHFTPDVCOILLDQSLDLAEYNFLPALSTTPTPDEVAAPSGFTLLATVNV 1620  
DB 1561 ILSTKLALRHFTPDVCOILLDQSLDLAEYNFLPALSTTPTPDEVAAPSGFTLLATVNV 1620  
QY 1621 ALNMLGELDKKKEPLTOAVGLSTQAEGRITLKSLLMFMEVCFYLLIQANRYLADPAVH 1680  
DB 1621 ALNMLGELDKKKEPLTOAVGLSTQAEGRITLKSLLMFMEVCFYLLIQANRYLADPAVH 1680  
QY 1681 PRDKORMQOELSELSTLSSLSRYFRGAPSSPATGVLPSPQKSTLSKASPESSQPL 1740  
DB 1681 PRDKORMQOELSELSTLSSLSRYFRGAPSSPATGVLPSPQKSTLSKASPESSQPL 1740  
QY 1741 IQLVOAFYRHMOR 1753  
DB 1741 IQLVOAFYRHMOR 1753

RESULT 6  
US-10-719-385-13  
; Sequence 13, Application US/10719385  
; Publication No. US20040209284A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Toole et al.  
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
; FILE REFERENCE: 22058-582  
; CURRENT FILING DATE: 2003-11-21  
; PRIOR FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: PCT/US03/37339  
; PRIOR FILING DATE: 2002-11-21  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 1753  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-719-385-13  
Query Match 100.0%; Score 9004; DB 17; Length 1753;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MIRSKITTSVLSFCRSSREMLTILGRSALRELSQIBALEMKWRRLLEGISTYKPPSPS 60  
DB 1 MIRSKITTSVLSFCRSSREMLTILGRSALRELSQIBALEMKWRRLLEGISTYKPPSPS 60  
QY 61 SAEVYKANKDVASPLKELGRLSKFGLGDEBSQVLOLCYLOENRYGRBDSVKTVLOER 120  
DB 61 SAEVYKANKDVASPLKELGRLSKFGLGDEBSQVLOLCYLOENRYGRBDSVKTVLOER 120  
QY 121 QSOALLIKIADYEEERTCILRCVHLITYPQDERHRYREYADCVKLEKELVSKYRQ 180  
DB 121 QSOALLIKIADYEEERTCILRCVHLITYPQDERHRYREYADCVKLEKELVSKYRQ 180  
QY 181 FEELYKTEAPWETHNGULMTERQVSRWFVQCLARQSMLEIIFLYAYFEMABSDLVLT 240  
DB 181 FEELYKTEAPWETHNGULMTERQVSRWFVQCLARQSMLEIIFLYAYFEMABSDLVLT 240  
QY 241 KMFKEQFGSRQTRHLYDETMDFVDRIGVFSALLIVEGMDISLHKCALDDREHLOF 300  
DB 241 KMFKEQFGSRQTRHLYDETMDFVDRIGVFSALLIVEGMDISLHKCALDDREHLOF 300  
QY 301 AODGLICQDMDCMLTFGDIPIHAPVLLAWALLRHTLNPETSSVVERIGGTALQLNVFO 360  
DB 301 AODGLICQDMDCMLTFGDIPIHAPVLLAWALLRHTLNPETSSVVERIGGTALQLNVFO 360  
QY 361 YLTRLLQSLASGNDCTTSTRCKVCYGLSFVLTSLBLHTGANOQDITDTRCEVLADBSL 420  
DB 361 YLTRLLQSLASGNDCTTSTRCKVCYGLSFVLTSLBLHTGANOQDITDTRCEVLADBSL 420  
QY 421 PELFWGTEPNSGLIITLDSVCGMPHLLSPLOLRALVSKSRKAKVYSLDMSFVNE 480  
DB 421 PELFWGTEPNSGLIITLDSVCGMPHLLSPLOLRALVSKSRKAKVYSLDMSFVNE 480  
QY 481 LYKHKPDIIVSHBEGTLMRRQTPKLLYPLIGQOTMARIPOGTGVQVMDLDRAYLVWMEYSY 540  
DB 481 LYKHKPDIIVSHBEGTLMRRQTPKLLYPLIGQOTMARIPOGTGVQVMDLDRAYLVWMEYSY 540  
QY 541 SSWTLFCEIBMLHVVSTADVIOHCORVXPIDLVHVKVISTDISIADCLLPITSRIYML 600  
DB 541 SSWTLFCEIBMLHVVSTADVIOHCORVXPIDLVHVKVISTDISIADCLLPITSRIYML 600  
QY 601 LORLTIVISPVVDIVASCVNCLTVLAARNPKAWTDLRHGFLPFVAHPVSSLSOMISAE 660  
DB 601 LORLTIVISPVVDIVASCVNCLTVLAARNPKAWTDLRHGFLPFVAHPVSSLSOMISAE 660  
QY 661 GMANAGYGNLMLNSEOPEGEYGVITIAFLRLITTLVKQOLGSGTOSQGLVPCVMFYLKEMLP 720  
DB 661 GMANAGYGNLMLNSEOPEGEYGVITIAFLRLITTLVKQOLGSGTOSQGLVPCVMFYLKEMLP 720

Db	661	GNNAGCYGNILMNSQPOSEYGVTTAFLRLITTLVKGQUGSTQSGVLPCWVFLKEMLP	720
Qy	721	SYHKRVNSHGVEQIGCLIELIHAILNLCHEITDLSHSTESLOFLCISLAYTEAGOT	780
Db	721	SYHKRVNSHGVEQIGCLIELIHAILNLCHEITDLSHSTESLOFLCISLAYTEAGOT	780
Qy	781	VININGIGVDITDMMAPPSRSDGEGGQGGOLLKTYGLASVTNNYRLRPPBNVSP	840
Db	781	VININGIGVDITDMMAPPSRSDGEGGQGGOLLKTYGLASVTNNYRLRPPBNVSP	840
Qy	841	LEQALSOHGAGNNLIVLAKYIYKHDPALPRLAJOLIKRLATYAPMSVYACLGNDAAA	900
Db	841	LEQALSOHGAGNNLIVLAKYIYKHDPALPRLAJOLIKRLATYAPMSVYACLGNDAAA	900
Qy	901	IRDAFLTLTQSKIEMRMKIMWLEFLTYAVEOPGJIELFLNLBYKQSGDSKESLSGMW	960
Db	901	IRDAFLTLTQSKIEMRMKIMWLEFLTYAVEOPGJIELFLNLBYKQSGDSKESLSGMW	960
Qy	961	SCLHAYLELIDSOQODRYWCPELTHRAAIAFTAHALMODRDSAMVLATKPFEMNLTSP	1020
Db	961	SCLHAYLELIDSOQODRYWCPELTHRAAIAFTAHALMODRDSAMVLATKPFEMNLTSP	1020
Qy	1021	LEFGLTSPSESEBSEILBTICALIMKICIEIYTVVXGSLDQSLKOTLKKSIEKFPAYMS	1080
Db	1021	LEFGLTSPSESEBSEILBTICALIMKICIEIYTVVXGSLDQSLKOTLKKSIEKFPAYMS	1080
Qy	1081	GYVKSIAHYAVETESSCTSLLEYOMLVSAMEMLIIATTHADIMHLLDSVVRQLPFDY	1140
Db	1081	GYVKSIAHYAVETESSCTSLLEYOMLVSAMEMLIIATTHADIMHLLDSVVRQLPFDY	1140
Qy	1141	LDGTKALLVPASVNCRLGSKCKTLLILILRQWREIGSVEIIGPTEILBEGVLQADQ	1200
Db	1141	LDGTKALLVPASVNCRLGSKCKTLLILILRQWREIGSVEIIGPTEILBEGVLQADQ	1200
Qy	1201	QJMEKTKAKVSAFTTVLOMKEMKXSDIPOYSQVLAVNCEITLQEEVIALFPQTRISLWLG	1260
Db	1201	QJMEKTKAKVSAFTTVLOMKEMKXSDIPOYSQVLAVNCEITLQEEVIALFPQTRISLWLG	1260
Qy	1261	SATEDKDSMETDDCSRSRHRDQDVCYLGHLANELCEVEDGDSWLOVTRRLPIEPTL	1320
Db	1261	SATEDKDSMETDDCSRSRHRDQDVCYLGHLANELCEVEDGDSWLOVTRRLPIEPTL	1320
Qy	1321	LTITLEVSLRMQNLHFTBATLHLLTLTARTQGAAYAVAGIGTOSICLPILSVYQSLNIG	1380
Db	1321	LTITLEVSLRMQNLHFTBATLHLLTLTARTQGAAYAVAGIGTOSICLPILSVYQSLNIG	1380
Qy	1381	TAQTPSASRKSULDAPSPGIVYRLSMSLEOQLKTLRYNFLPEALDFVGVHQBERTLOCLNA	1440
Db	1381	TAQTPSASRKSULDAPSPGIVYRLSMSLEOQLKTLRYNFLPEALDFVGVHQBERTLOCLNA	1440
Qy	1441	VRTYQSLACLBEADHTVGFIIQLSNFMKEMHFLPOLMRDIQVNLGYCOACTSLHSRK	1500
Db	1441	VRTYQSLACLBEADHTVGFIIQLSNFMKEMHFLPOLMRDIQVNLGYCOACTSLHSRK	1500
Qy	1501	MLQHYLQKNGDGPSSAQAQVQPPSAAASAPSSSKOPADPTEASEQALHTYQYGLK	1560
Db	1501	MLQHYLQKNGDGPSSAQAQVQPPSAAASAPSSSKOPADPTEASEQALHTYQYGLK	1560
Qy	1561	ILSKTLAALRHFTPDVCOILIDQSLDAEYNFLPALSTFTTFPSEVAPSGTLLATVNV	1620
Db	1561	ILSKTLAALRHFTPDVCOILIDQSLDAEYNFLPALSTFTTFPSEVAPSGTLLATVNV	1620
Qy	1621	ALNNMLGELDKKKEPITQAVGISTQAGEGRITLSILMTFMENCIFYLLISQARBYRDPAYH	1680
Db	1621	ALNNMLGELDKKKEPITQAVGISTQAGEGRITLSILMTFMENCIFYLLISQARBYRDPAYH	1680
Qy	1681	PRDKQMKQELSELSTLSSLSRYFRGACASSPATVGLPPOGKSTISLASSESOEPL	1740
Db	1681	PRDKQMKQELSELSTLSSLSRYFRGACASSPATVGLPPOGKSTISLASSESOEPL	1740
Qy	1741	IQLVQAFVRRHQR 1753	
Db	1741	IQLVQAFVRRHQR 1753	

```

RESULT 7
US-10-719-385-14
/ Sequence 14, Application US/10719385
/ Publication No. US20040209284A1
/ GENERAL INFORMATION:
/ APPLICANT: O'Toole et al.
/ TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
/ FILE REFERENCE: 22058-582
/ CURRENT FILING DATE: 2003-11-21
/ PRIOR APPLICATION NUMBER: PCT/US03/37339
/ PRIOR FILING DATE: 2003-11-21
/ PRIOR APPLICATION NUMBER: 60/426,094
/ PRIOR FILING DATE: 2002-11-21
/ NUMBER OF SEQ ID NOS: 26
/ SOFTWARE: Patencin Ver. 2.1
/ SEQ ID NO 14
/ LENGTH: 1753
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-719-385-14

Query Match      100.0%; Score 9004; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRKSKITSVYSPFCSSSRELTITLLGRSALNELSOTAELELNKHWRLLEGLSYKPPSP 60
DB 1 MIRKSKITSVSPFCSSSRELTITLLGRSALNELSOTAELELNKHWRLLEGLSYKPPSP 60
QY 61 SAEKVKANKDVASPLKEELGRLISKFLGDEBSQVOLLQCYLOEDYKGRTRDSYKTVLQDER 120
DB 61 SAEKVKANKDVASPLKEELGRLISKFLGDEBSQVOLLQCYLOEDYKGRTRDSYKTVLQDER 120
QY 121 OSQALLIKIADYYVYEEHRTCIACVLIHLITTYODERHPRVAYADQVDCLEKELYSKYKQ 180
DB 121 OSQALLIKIADYYVYEEHRTCIACVLIHLITTYODERHPRVAYADQVDCLEKELYSKYKQ 180
QY 181 FEELYKTEAPWETHGNLMTERQVSRWVQCLREOSMLEIIFLYAAYFEMAPSDLVLT 240
DB 181 FEELYKTEAPWETHGNLMTERQVSRWVQCLREOSMLEIIFLYAAYFEMAPSDLVLT 240
QY 241 KMFKEQGFSGSQTRNHLVDETMDFVDRIGYFSALLIVEGMDISLHKCALDDRREHLQF 300
DB 241 KMFKEQGFSGSQTRNHLVDETMDFVDRIGYFSALLIVEGMDISLHKCALDDRREHLQF 300
QY 301 AODGLICODMCMILTRGDIPIHNAPLYLAWALLRHTINPETSIVYRKIGTATOLNVFO 360
DB 301 AODGLICODMCMILTRGDIPIHNAPLYLAWALLRHTINPETSIVYRKIGTATOLNVFO 360
QY 361 YLRLLOSLASGGNDCTTSTACMCVYGLISFVLTSLEHLTGNODDITDTACEVLADPSL 420
DB 361 YLRLLOSLASGGNDCTTSTACMCVYGLISFVLTSLEHLTGNODDITDTACEVLADPSL 420
QY 421 PELFWGTHPEPISGLIILDSVCGMPHLSPLLOLRALVSGSKTRAKKYSLDKMSFVNE 480
DB 421 PELFWGTHPEPISGLIILDSVCGMPHLSPLLOLRALVSGSKTRAKKYSLDKMSFVNE 480
QY 481 LYKXKHPHDVISHEGTLMRRQTPKLLVPLGGQTMARIPOGVGQVMDLDRAYLVKMEVSY 540
DB 481 LYKXKHPHDVISHEGTLMRRQTPKLLVPLGGQTMARIPOGVGQVMDLDRAYLVKMEVSY 540
QY 541 SSWTLFTCEIEMLLHVSTADVIHQCRVKPDIIDLVKVISTDLSIADCLPITSRIYML 600
DB 541 SSWTLFTCEIEMLLHVSTADVIHQCRVKPDIIDLVKVISTDLSIADCLPITSRIYML 600
QY 601 LQRLITVISPVDVIAACVNCITVLARNAPKWTDLRHGFLPFVAHPVSSLSQMSAE 660
DB 601 LQRLITVISPVDVIAACVNCITVLARNAPKWTDLRHGFLPFVAHPVSSLSQMSAE 660
QY 661 GMAAGGVGNILMNSEQEGEGYVIAELRLITTLVKGOLGSGTQSGVLPVCMFVLKEMLP 720

```



```

Db 661 GMAAGYGNLMSNQPGEGVTIAFLRLITTVLKGGLGSGVGLPCMFVLKEMLP 720
Qy 721 SYHKRWNSHVRBOIGCLIELLHAIINLCHETDLSHSHPSLOFLCISLAYEAGOT 780
Db 721 SYHKRWNSHVRBOIGCLIELLHAIINLCHETDLSHSHPSLOFLCISLAYEAGOT 780
Qy 781 VINIMGIVDTIDWMAQPSDAGEGQGLIKTKYLA.PSVYNNVYRLKPPSNVSP 840
Db 781 VINIMGIVDTIDWMAQPSDAGEGQGLIKTKYLA.PSVYNNVYRLKPPSNVSP 840
Qy 841 LEQALSQGHAGNNLIATLAKYIYKHDPALPRLAIQLLKLATVAPMSVYACLGNDAAA 900
Db 841 LEQALSQGHAGNNLIATLAKYIYKHDPALPRLAIQLLKLATVAPMSVYACLGNDAAA 900
Qy 901 IRDAFLTRLQSKIEPMRIKMWLEFLYAVETQPELIFLNLBYKQSDSDSKESLGMW 960
Db 901 IRDAFLTRLQSKIEPMRIKMWLEFLYAVETQPELIFLNLBYKQSDSDSKESLGMW 960
Qy 961 SCLHNVLELIDSOQODRYWCPPLHRAAIAFLHLMODRDSAMLVLRTPKPMENLTS 1020
Db 961 SCLHNVLELIDSOQODRYWCPPLHRAAIAFLHLMODRDSAMLVLRTPKPMENLTS 1020
Qy 1021 LFGTLPSPSETSEPSILETCALIMKICLEIYVVVKSJLQSLKXDTLKKFSIEKRFAYMS 1080
Db 1021 LFGTLPSPSETSEPSILETCALIMKICLEIYVVVKSJLQSLKXDTLKKFSIEKRFAYMS 1080
Qy 1081 GYVKSVAHVAVETBSSCTSLLEYOMLVANRMLIITTHADIMHLDTSVVRQLFDV 1140
Db 1081 GYVKSVAHVAVETBSSCTSLLEYOMLVANRMLIITTHADIMHLDTSVVRQLFDV 1140
Qy 1141 LDGTRALLVPASVNCRLGSMKCTLLILRLKQRELSVDEILGPTLELLEGLQADQ 1200
Db 1141 LDGTRALLVPASVNCRLGSMKCTLLILRLKQRELSVDEILGPTLELLEGLQADQ 1200
Qy 1201 QLMEXTKAVESAFITVLQMKEMKVS.DI.PQYSQVLVANCETLQOEVIALPQTRHSLALG 1260
Db 1201 QLMEXTKAVESAFITVLQMKEMKVS.DI.PQYSQVLVANCETLQOEVIALPQTRHSLALG 1260
Qy 1261 SATEDKDSMETDDCSRHRRDQDQVYGLHLAKELCEVDEDDGSMQVTRRLPILPTL 1320
Db 1261 SATEDKDSMETDDCSRHRRDQDQVYGLHLAKELCEVDEDDGSMQVTRRLPILPTL 1320
Qy 1321 LTTLEVSIRMKONLHFTBATLHLLTLARTQOGAFAVAGAGITOSICPLSVYQLSTNG 1380
Db 1321 LTTLEVSIRMKONLHFTBATLHLLTLARTQOGAFAVAGAGITOSICPLSVYQLSTNG 1380
Qy 1381 TAQTPSASRKSJLDA.PSWPGVYRLSMSLMBOQLKTLRYNFLPEALDFVGHQERTLOCLNA 1440
Db 1381 TAQTPSASRKSJLDA.PSWPGVYRLSMSLMBOQLKTLRYNFLPEALDFVGHQERTLOCLNA 1440
Qy 1441 VRTVQSLACLEBADHTVGFIIQLSNFMKEMPHFLPOLMRDIOVNGYVCOACTSLHSHRK 1500
Db 1441 VRTVQSLACLEBADHTVGFIIQLSNFMKEMPHFLPOLMRDIOVNGYVCOACTSLHSHRK 1500
Qy 1501 MLQHYLQNKNDGJLPSA.VAORVORPSPASAPSSSKOPADPTESEGOALHTVQYGLK 1560
Db 1501 MLQHYLQNKNDGJLPSA.VAORVORPSPASAPSSSKOPADPTESEGOALHTVQYGLK 1560
Qy 1561 ILSTKLALRHFTPDVCOILLDQSLDLAEYNFLPALSTFTPTPDESEVAPSGTLLATVNV 1620
Db 1561 ILSTKLALRHFTPDVCOILLDQSLDLAEYNFLPALSTFTPTPDESEVAPSGTLLATVNV 1620
Qy 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRITLKSJLMFTMENCYLLIQAMRYLRDPVH 1680
Db 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRITLKSJLMFTMENCYLLIQAMRYLRDPVH 1680
Qy 1681 PRDKORMKOELSSSELSTLSLSRFRFGAPSSPATVGLPSPOGKSTLSKASPSQBP 1740
Db 1681 PRDKORMKOELSSSELSTLSLSRFRFGAPSSPATVGLPSPOGKSTLSKASPSQBP 1740
Qy 1741 IOLVOAFVRHMQR 1753

```

```

Db 1741 IOLVOAFVRHMQR 1753
RESULT 8
US-10-719-385-15
; Sequence 15, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-15
Query Match 100.0%; Score 9003; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1751; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MIRSKITSVLSFCRSSRELTIIILGRSALRELSQIEALNKHWRLLLEGLSYKPPSPS 60
Db 1 MIRSKITSVLSFCRSSRELTIIILGRSALRELSQIEALNKHWRLLLEGLSYKPPSPS 60
Qy 61 SAEYKANKDVASPLKELGLRISKFLGIDESQSVQLLOCTYQBYRGRDSVKTVLQDER 120
Db 61 SAEYKANKDVASPLKELGLRISKFLGIDESQSVQLLOCTYQBYRGRDSVKTVLQDER 120
Qy 121 OSQALIKIADYVEERTCIACVLAHLTYRQDRHPRVAYACVDLKEELYSKTRQ 180
Db 121 OSQALIKIADYVEERTCIACVLAHLTYRQDRHPRVAYACVDLKEELYSKTRQ 180
Qy 181 FEELYKTEAPTWETHGNLMTERQVSRWFVQCLREOSMLEIIFLYAVFEMAPSDLVLT 240
Db 181 FEELYKTEAPTWETHGNLMTERQVSRWFVQCLREOSMLEIIFLYAVFEMAPSDLVLT 240
Qy 241 KMFYKQGFSGQTRNHLVDETMDFVDRIGYFSALLIVEGMDISLHKCALDDRRRLHQF 300
Db 241 KMFYKQGFSGQTRNHLVDETMDFVDRIGYFSALLIVEGMDISLHKCALDDRRRLHQF 300
Qy 301 AODGLIQDMDCLMTFGDIPHNAPVLLAMALRHHTLNPEETSIVRKIGGTAIQLANVQ 360
Db 301 AODGLIQDMDCLMTFGDIPHNAPVLLAMALRHHTLNPEETSIVRKIGGTAIQLANVQ 360
Qy 361 YLTRLLQSLASGANDCTTSTACMCVYGLSFLVLSLELHTLGNQODIIDTA.CEVLADPSL 420
Db 361 YLTRLLQSLASGANDCTTSTACMCVYGLSFLVLSLELHTLGNQODIIDTA.CEVLADPSL 420
Qy 421 PELFWGTEPTSGIGIILDSVCGMPHILSP.LQLRALVSGKSTRAKTYSTLDDKSFYNE 480
Db 421 PELFWGTEPTSGIGIILDSVCGMPHILSP.LQLRALVSGKSTRAKTYSTLDDKSFYNE 480
Qy 481 LYKXKPHDVISHEDGTLMRQTPKLLVPLGGQTMURIPQGVGVMLDDBRAYLRWESY 540
Db 481 LYKXKPHDVISHEDGTLMRQTPKLLVPLGGQTMURIPQGVGVMLDDBRAYLRWESY 540
Qy 541 SSWTLFTCEIMELHVVSTADVIOHQKRVKEIIDLVHKVSTJDLSTADCLLPITSRIYML 600
Db 541 SSWTLFTCEIMELHVVSTADVIOHQKRVKEIIDLVHKVSTJDLSTADCLLPITSRIYML 600
Qy 601 LQRLITTVISPPVDVIASCVNCLTVLAARNPAKWTDLRHGTFLFPVAHPVSSLSQMSAE 660
Db 601 LQRLITTVISPPVDVIASCVNCLTVLAARNPAKWTDLRHGTFLFPVAHPVSSLSQMSAE 660

```



```

Qy 661 GNNAGGYGNILMNSBOPGEYGTIAFLRLITTLVKGOLSGTOSQGLVPCVWFVXKMLP 720
Db 661 GNNAGGYGNILMNSBOPGEYGTIAFLRLITTLVKGOLSGTOSQGLVPCVWFVXKMLP 720
Qy 721 SYHKRRYNSHGVREIOGCLILELIIAIIIMLCHETDLSSHPISLOFLCISIAYTEAOQT 780
Db 721 SYHKRRYNSHGVREIOGCLILELIIAIIIMLCHETDLSSHPISLOFLCISIAYTEAOQT 780
Qy 781 VININGIGVDTIDMMAAOPRSDGABGGQGGOLLKTYKLAFSYNNNTIRLKPBNVYSP 840
Db 781 VININGIGVDTIDMMAAOPRSDGABGGQGGOLLKTYKLAFSYNNNTIRLKPBNVYSP 840
Qy 841 LEQALSOHAGHNNILAVLAKYIYKHDPALPRLAIQILKRLATVAPMSVYACIGNDAAA 900
Db 841 LEQALSOHAGHNNILAVLAKYIYKHDPALPRLAIQILKRLATVAPMSVYACIGNDAAA 900
Qy 901 IRDALVTLRQSKIEEMRIKMWLELVYAVETQPCLEFLNLEYKDSGDSKESFLGMV 960
Db 901 IRDALVTLRQSKIEEMRIKMWLELVYAVETQPCLEFLNLEYKDSGDSKESFLGMV 960
Qy 961 SCIAHVLILDSQQODRYWCPPLHRAAIAFLHLMODRDSAMVLRTKPFMENLTSF 1020
Db 961 SCIAHVLILDSQQODRYWCPPLHRAAIAFLHLMODRDSAMVLRTKPFMENLTSF 1020
Qy 1021 LFGTLPSPSETSEPSILETCALIMKICLEIYVVVKGSLDQSLKDTLKFSIEKRFAYMS 1080
Db 1021 LFGTLPSPSETSEPSILETCALIMKICLEIYVVVKGSLDQSLKDTLKFSIEKRFAYMS 1080
Qy 1081 GYVSLAHVAETESSCTSLIEYOMLVSAWPMILITTHADIMHLDTSVVRQOLFV 1140
Db 1081 GYVSLAHVAETESSCTSLIEYOMLVSAWPMILITTHADIMHLDTSVVRQOLFV 1140
Qy 1141 LDGTALLLVASVNCILGSMKCTLLILIRQWREIGSVDEILGPULTEILEGVLQADQ 1200
Db 1141 LDGTALLLVASVNCILGSMKCTLLILIRQWREIGSVDEILGPULTEILEGVLQADQ 1200
Qy 1201 QLMETKXKVSAPFTTVQMKEMKVSIDPOYSQVLANCETLOEYVALPQTRHSIALG 1260
Db 1201 QLMETKXKVSAPFTTVQMKEMKVSIDPOYSQVLANCETLOEYVALPQTRHSIALG 1260
Qy 1261 SATEDKDSMETDSCRSRHRDOROVGYLGHAKELCEVDEDDGSMLOVTRRLPIPLTL 1320
Db 1261 SATEDKDSMETDSCRSRHRDOROVGYLGHAKELCEVDEDDGSMLOVTRRLPIPLTL 1320
Qy 1321 LTTLEVSIRMKONTLHTEATLHLLTLARTOQATAVAGAGITQSCILPLSVYQLSTNG 1380
Db 1321 LTTLEVSIRMKONTLHTEATLHLLTLARTOQATAVAGAGITQSCILPLSVYQLSTNG 1380
Qy 1381 TAQTPSASRSKSLDAPSWGVYRLSMSLMEOLKTLRYNFPLEADPFGVHOERTLOCLNA 1440
Db 1381 TAQTPSASRSKSLDAPSWGVYRLSMSLMEOLKTLRYNFPLEADPFGVHOERTLOCLNA 1440
Qy 1441 VRTVQSLACLEBAHTVGFILQLSNFMKEWHFHLPOLMRDIQVNLGYLCOACTSLHSRK 1500
Db 1441 VRTVQSLACLEBAHTVGFILQLSNFMKEWHFHLPOLMRDIQVNLGYLCOACTSLHSRK 1500
Qy 1501 MLOHYLONKNDGJPSAVAQRYQRPSSASAAPSSSQPADTEASEQALHTVQYGLK 1560
Db 1501 MLOHYLONKNDGJPSAVAQRYQRPSSASAAPSSSQPADTEASEQALHTVQYGLK 1560
Qy 1561 ILSTLALAHFTEPVOOILLDOSLDLAENVLPALSFPTTPRSEVAPASGTLAVNV 1620
Db 1561 ILSTLALAHFTEPVOOILLDOSLDLAENVLPALSFPTTPRSEVAPASGTLAVNV 1620
Qy 1621 ALNMLGELDKKEPLTQAVGLSTQAEGRITLSLMTMENCYLLISQANRYLRDPAVH 1680
Db 1621 ALNMLGELDKKEPLTQAVGLSTQAEGRITLSLMTMENCYLLISQANRYLRDPAVH 1680
Qy 1681 PRDKQKMKOISLSTLSTLSRYFRGAPSSPATGVLPSPQCKSTSLSKASPESEGPL 1740
Db 1681 PRDKQKMKOISLSTLSTLSRYFRGAPSSPATGVLPSPQCKSTSLSKASPESEGPL 1740
Qy 1741 IQLVQAFVRHMQR 1753

```

```

Db 1741 IQLVQAFVRHMQR 1753

RESULT 9
US-10-719-385-6
; Sequence 6, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-6

```

```

Query Match 99.9%; Score 9002; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1753; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MIRSKITSVLSFCRSSREIMTILGRSALRELSQIEAEINKMRRLLBGLSYKKPPSPS 60
Db 1 MIRSKITSVLSFCRSSREIMTILGRSALRELSQIEAEINKMRRLLBGLSYKKPPSPS 60
Qy 61 SAEKVANKVAPPLKRLGIRISKEPLGLDEBSVOLLOCYLQEDYGRDTSVKTVALDER 120
Db 61 SAEKVANKVAPPLKRLGIRISKEPLGLDEBSVOLLOCYLQEDYGRDTSVKTVALDER 120
Qy 121 QSOALILKIDVYEEETCIRCVLHLITVFODERHPYREYADCVDKELVSKYRQ 180
Db 121 QSOALILKIDVYEEETCIRCVLHLITVFODERHPYREYADCVDKELVSKYRQ 180
Qy 181 FEEIYKTEAPWETHGNLMTERRQVSRWFVQCLREQSMLEIIFLYYAYEMAPSDDLVL 240
Db 181 FEEIYKTEAPWETHGNLMTERRQVSRWFVQCLREQSMLEIIFLYYAYEMAPSDDLVL 240
Qy 241 KMKKEGFGSRQNRRLNDETMDPFDVRIQYPSALILVSGMDIESLHKCALDRRELHOF 300
Db 241 KMKKEGFGSRQNRRLNDETMDPFDVRIQYPSALILVSGMDIESLHKCALDRRELHOF 300
Qy 301 AODGLICODMDCLMTFGDI PHHA PVLAMALLRHTLANBETSSVVRKIGGTAIQNLVFC 360
Db 301 AODGLICODMDCLMTFGDI PHHA PVLAMALLRHTLANBETSSVVRKIGGTAIQNLVFC 360
Qy 361 YLTRLQSLASGGNDCTTSTACVYGLISFVLTSLHHTLGNQODI IDTACEVLADPSL 420
Db 361 YLTRLQSLASGGNDCTTSTACVYGLISFVLTSLHHTLGNQODI IDTACEVLADPSL 420
Qy 421 PELFMGTPEPSGILILDSVCGMFPPLSLPIQLALVSGSTAKVYSPDKMFPYNE 480
Db 421 PELFMGTPEPSGILILDSVCGMFPPLSLPIQLALVSGSTAKVYSPDKMFPYNE 480
Qy 481 LYHGRHDVISHEDGTLMRROT PKLXPLPGGQTNLRI PGTVGQVMLDDRVLVREYSY 540
Db 481 LYHGRHDVISHEDGTLMRROT PKLXPLPGGQTNLRI PGTVGQVMLDDRVLVREYSY 540
Qy 541 SSWTLFTCEIEMLHVASTADYI OHQORVYPIIDLVHKYISTDLSTADCLLPITSRYWL 600
Db 541 SSWTLFTCEIEMLHVASTADYI OHQORVYPIIDLVHKYISTDLSTADCLLPITSRYWL 600
Qy 601 LORLTVIAPPVDVASCNCLTVLAARNPAKVTMDLHRTGFLPFAHVSSLSQWISAE 660
Db 601 LORLTVIAPPVDVASCNCLTVLAARNPAKVTMDLHRTGFLPFAHVSSLSQWISAE 660

```

QY 661 GNNAGYGNILMNSBQCEGYVTIARLITTLVKQGLSTOSQGLVPCWFKLXELP 720  
DB 661 GNNAGYGNILMNSBQCEGYVTIARLITTLVKQGLSTOSQGLVPCWFKLXELP 720  
QY 721 SYHKRYNSHGVREIGLILIELIHAITLCHETDLSHSPSLQFLCTISLATEAGOT 780  
DB 721 SYHKRYNSHGVREIGLILIELIHAITLCHETDLSHSPSLQFLCTISLATEAGOT 780  
QY 781 VINIMIGVDTIIDWMAAOPRSDGEGGQGLIKTYKLAFSVTNNVIRLKPSSNVSP 840  
DB 781 VINIMIGVDTIIDWMAAOPRSDGEGGQGLIKTYKLAFSVTNNVIRLKPSSNVSP 840  
QY 841 LEQALSQHGAGNNLIATLAKYIYHKDPAIPRLAIQILKRLATVAPMSVYACGNDAAA 900  
DB 841 LEQALSQHGAGNNLIATLAKYIYHKDPAIPRLAIQILKRLATVAPMSVYACGNDAAA 900  
QY 901 IRDAFLTRLOSKIEDMRKMWLEFLVAVEQPOLIELFNLEVKDSDGSKESFLGAM 960  
DB 901 IRDAFLTRLOSKIEDMRKMWLEFLVAVEQPOLIELFNLEVKDSDGSKESFLGAM 960  
QY 961 SCLHAVLELIDSOQDRYWCPEPLHRAAIAFLHLMQDRDSAMLVLRKPKFMENTTSP 1020  
DB 961 SCLHAVLELIDSOQDRYWCPEPLHRAAIAFLHLMQDRDSAMLVLRKPKFMENTTSP 1020  
QY 1021 LFGTLSPSETSEBSILETCALIMKIICLETIYVYVKGSLDSDLKDTLKKESTIEKRFAPWS 1080  
DB 1021 LFGTLSPSETSEBSILETCALIMKIICLETIYVYVKGSLDSDLKDTLKKESTIEKRFAPWS 1080  
QY 1081 GYVSLAVHVAETGSSCTSLLEYOMLVSAWRMLIITTHADIMHLLDSVVRQOLFIDV 1140  
DB 1081 GYVSLAVHVAETGSSCTSLLEYOMLVSAWRMLIITTHADIMHLLDSVVRQOLFIDV 1140  
QY 1141 LDGTRALLLVASVNCRLGSMKCTLLIILRQWRKELGSDVEILIGPLTEILLEGVLQADQ 1200  
DB 1141 LDGTRALLLVASVNCRLGSMKCTLLIILRQWRKELGSDVEILIGPLTEILLEGVLQADQ 1200  
QY 1201 QLMKRTAKVSAFTIYQMKEMKVSDFIYQSOLVANCETLOEBVIALPOTRHSALG 1260  
DB 1201 QLMKRTAKVSAFTIYQMKEMKVSDFIYQSOLVANCETLOEBVIALPOTRHSALG 1260  
QY 1261 SATEDKDSMETDSCSRSHRQDQVCYLGHLAKELCEVEDSDSMQVTRRLPIEPTL 1320  
DB 1261 SATEDKDSMETDSCSRSHRQDQVCYLGHLAKELCEVEDSDSMQVTRRLPIEPTL 1320  
QY 1321 LTTLEVSIRMKONLHFTBATLHLITLARTOQATAVAGAGITOSICLPILSVYQLSTNG 1380  
DB 1321 LTTLEVSIRMKONLHFTBATLHLITLARTOQATAVAGAGITOSICLPILSVYQLSTNG 1380  
QY 1381 TAQTPSASRSKSLDAPSWGVYRLSLSMEOLIKTRNFLEPBLDFVGVHQBRTLOCLANA 1440  
DB 1381 TAQTPSASRSKSLDAPSWGVYRLSLSMEOLIKTRNFLEPBLDFVGVHQBRTLOCLANA 1440  
QY 1441 VRTVQSLACEADHTVGFIIOLSNFMKEMPHLPQLMRDIOVNLGYCOACTSLHSRK 1500  
DB 1441 VRTVQSLACEADHTVGFIIOLSNFMKEMPHLPQLMRDIOVNLGYCOACTSLHSRK 1500  
QY 1501 MLQHYLQNKDGLPSAQAQVORPPSAASAAPSSSKOPADTEASEQOALHTVQYGLK 1560  
DB 1501 MLQHYLQNKDGLPSAQAQVORPPSAASAAPSSSKOPADTEASEQOALHTVQYGLK 1560  
QY 1561 ILSTKTLAALRHFTPDVCOILLDOSJDLAEVNFPLALSTTPTFSEVAPSGTLLATYV 1620  
DB 1561 ILSTKTLAALRHFTPDVCOILLDOSJDLAEVNFPLALSTTPTFSEVAPSGTLLATYV 1620  
QY 1621 ALNMLGELDKKKEPLTOAVGISTOAGRTLKSILMFTMENC FYLLISQARVYRDPVH 1680  
DB 1621 ALNMLGELDKKKEPLTOAVGISTOAGRTLKSILMFTMENC FYLLISQARVYRDPVH 1680  
QY 1681 PRDKORMKQBLSELSTLSSLSRYFRGAPSSPATVGLPSPQKSTSLASAPESQBL 1740  
DB 1681 PRDKORMKQBLSELSTLSSLSRYFRGAPSSPATVGLPSPQKSTSLASAPESQBL 1740

QY 1741 IQLVQAFVRHMOR 1753  
DB 1741 IQLVQAFVRHMOR 1753  
RESULT 10  
US-10-719-385-10  
; Sequence 10, Application US/10719385  
; Publication No. US20040209284A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Toole et al.  
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
; FILE REFERENCE: 22058-582  
; CURRENT APPLICATION NUMBER: US/10/719,385  
; PRIOR FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: PCT/US03/37339  
; PRIOR FILING DATE: 2002-11-21  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 1753  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-719-385-10  
Query Match 99.9%; Score 9002; DB 17; Length 1753;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITTSVLSFRSSRELITILLGSALRELSQIEAELNKHRRILLEGISYKPPSPS 60  
DB 1 MIRSKITTSVLSFRSSRELITILLGSALRELSQIEAELNKHRRILLEGISYKPPSPS 60  
QY 61 SAEKYKANKQVASPLKELGLRISKFLGIDEEQVOLLQCYLOEDYRGTRDSVKTVALDER 120  
DB 61 SAEKYKANKQVASPLKELGLRISKFLGIDEEQVOLLQCYLOEDYRGTRDSVKTVALDER 120  
QY 121 QSQALLIKIADYYEERTCIIRCVLHLITFYFODERHPYRVAYADCVDLKEVELSKTRQ 180  
DB 121 QSQALLIKIADYYEERTCIIRCVLHLITFYFODERHPYRVAYADCVDLKEVELSKTRQ 180  
QY 181 FEELYKTEAPWETHNGULMNEROYSRWVQCLREOSMLLETIFYVAYFEAPSDLVLT 240  
DB 181 FEELYKTEAPWETHNGULMNEROYSRWVQCLREOSMLLETIFYVAYFEAPSDLVLT 240  
QY 241 KMFKEQGFSGRQTRHLYVDETMDPFVDRIGYFSALLIVEGMDISLHKCALDDREBLHQF 300  
DB 241 KMFKEQGFSGRQTRHLYVDETMDPFVDRIGYFSALLIVEGMDISLHKCALDDREBLHQF 300  
QY 301 AODGLICQDMDCMLTFRGDIIPHNAVLLAMALLRHTINPEETSSVVRKIGGTAIQLNVQ 360  
DB 301 AODGLICQDMDCMLTFRGDIIPHNAVLLAMALLRHTINPEETSSVVRKIGGTAIQLNVQ 360  
QY 361 YLTRLOSLASGNDCTTSTKCMCVYGLLSVLTSLBHTIGNOODIITPACVTLADPSL 420  
DB 361 YLTRLOSLASGNDCTTSTKCMCVYGLLSVLTSLBHTIGNOODIITPACVTLADPSL 420  
QY 421 PELFMGERTPSGLIILSDVCGMPPHLLSPLOLRLALVSGKSTAKKYVSLDKMSFYNE 480  
DB 421 PELFMGERTPSGLIILSDVCGMPPHLLSPLOLRLALVSGKSTAKKYVSLDKMSFYNE 480  
QY 481 LYKHKPDVISHEDGTLMRQTPRLVPLGGQTNLRIPOGTVGQVMLDRAVLYVRWESY 540  
DB 481 LYKHKPDVISHEDGTLMRQTPRLVPLGGQTNLRIPOGTVGQVMLDRAVLYVRWESY 540  
QY 541 SSWTLFCEIEMLHVSTADVIOHCORVXIIDLVHKVISTDLSADCLLPITSRIYML 600  
DB 541 SSWTLFCEIEMLHVSTADVIOHCORVXIIDLVHKVISTDLSADCLLPITSRIYML 600  
QY 601 LQRLTIVTSPVDVIASCVNCLTVLAARNPAKVTMDLHRTGFLPFVAHPVSSLQMTSAB 660

601 LORLTVISPPVDVIAACVNCCLTVLAARNPAKWTDLRHGTGFLPVVAHPVSSLSQMSIAE 660  
661 GNNAGGYGNLNNSEPOGEYGVTTIAFLRLITLVKSGLSQSGVLPCWPFVUKEMLP 720  
661 GNNAGGYGNLNNSEPOGEYGVTTIAFLRLITLVKSGLSQSGVLPCWPFVUKEMLP 720  
721 SYHKRNVSHGVREOIGCLILEL1HA1NMLCHETDLHSHSTPSLOFLCISLATEAGOT 780  
721 SYHKRNVSHGVREOIGCLILEL1HA1NMLCHETDLHSHSTPSLOFLCISLATEAGOT 780  
721 VININGIGVDTIDMMAAOPRSDGAGGCGGGLIKTYKLAFTSVNNVIRLKPSPNVSP 840  
721 VININGIGVDTIDMMAAOPRSDGAGGCGGGLIKTYKLAFTSVNNVIRLKPSPNVSP 840  
841 LEQALSOHAGNNMLIAVLAKYIYKHGPALPRLAIOQLKRLATAPAPSVACIGNDAAA 900  
841 LEQALSOHAGNNMLIAVLAKYIYKHGPALPRLAIOQLKRLATAPAPSVACIGNDAAA 900  
901 IRDAFLTLQSLKIEEMRIKVMILEFLTVAVETOPGLIELFNLNLEKSDGSKESPLGMW 960  
901 IRDAFLTLQSLKIEEMRIKVMILEFLTVAVETOPGLIELFNLNLEKSDGSKESPLGMW 960  
961 SC1HAYLELIDSOQODRYWCPPLHRAAIAFLHLMODRDSAMLVYTRKPEMENTLSP 1020  
961 SC1HAYLELIDSOQODRYWCPPLHRAAIAFLHLMODRDSAMLVYTRKPEMENTLSP 1020  
1021 LFGTLPSPSETSEPSILETCALIMKICLIEIYVVVKSIDLSKDTLKKFSIERKPAWMS 1080  
1021 LFGTLPSPSETSEPSILETCALIMKICLIEIYVVVKSIDLSKDTLKKFSIERKPAWMS 1080  
1081 GYVSLAHVAETESSCTSLIEYOMLVSAWMLIITTHADIMHLDTSVVRQQLFDV 1140  
1081 GYVSLAHVAETESSCTSLIEYOMLVSAWMLIITTHADIMHLDTSVVRQQLFDV 1140  
1141 LDGTAKALLVPAVNCLEIGSMKCTLALLILRQWRELSYDILGPTTEILEBVLAQDQ 1200  
1141 LDGTAKALLVPAVNCLEIGSMKCTLALLILRQWRELSYDILGPTTEILEBVLAQDQ 1200  
1201 QLMERTYAKVSAFTTVLQMKEMKVSIDIPQYSQVLVNCETLQEBVIALFDQTRHSIATG 1260  
1201 QLMERTYAKVSAFTTVLQMKEMKVSIDIPQYSQVLVNCETLQEBVIALFDQTRHSIATG 1260  
1261 SATEDKSMETDSCSRSHRQDQVCLGLHAKELCEVEDDSDMLQVTRRLPIPTL 1320  
1261 SATEDKSMETDSCSRSHRQDQVCLGLHAKELCEVEDDSDMLQVTRRLPIPTL 1320  
1321 LTTLEVSILRMKONLHFEATLHLTLARTQOGATAVAGAGITOSICLPILSVYQJSTNG 1380  
1321 LTTLEVSILRMKONLHFEATLHLTLARTQOGATAVAGAGITOSICLPILSVYQJSTNG 1380  
1381 TAQTPSASRSKSLDAPSPGCVRLSMSLMEQLKTLARNFLPEALDFGVHQBERTLOCLNA 1440  
1381 TAQTPSASRSKSLDAPSPGCVRLSMSLMEQLKTLARNFLPEALDFGVHQBERTLOCLNA 1440  
1441 VRTVQSLACLEADHTVGFILQSLNPKMEWHFHLPOLMRDIOVNLGYLCOACTSLHSRK 1500  
1441 VRTVQSLACLEADHTVGFILQSLNPKMEWHFHLPOLMRDIOVNLGYLCOACTSLHSRK 1500  
1501 MLQHTLONKKGDGIPSAVAORVORPSPAASAPSSSROPADTEASBOALHTVOYGLK 1560  
1501 MLQHTLONKKGDGIPSAVAORVORPSPAASAPSSSROPADTEASBOALHTVOYGLK 1560  
1561 ILSTKTLAALHFFPDDVOOILLDOSLDLAENFLPALSFPTPTDSEVAPBFGTLATVNV 1620  
1561 ILSTKTLAALHFFPDDVOOILLDOSLDLAENFLPALSFPTPTDSEVAPBFGTLATVNV 1620  
1621 ALNMLGELDKKKEFLTOAVGLSTQABGRTLKSLMTMENCYLLISQAMRYLDPAYH 1680  
1621 ALNMLGELDKKKEFLTOAVGLSTQABGRTLKSLMTMENCYLLISQAMRYLDPAYH 1680  
1681 PRDKOKMOKELSELSTLSSLSRYFRGAPSSPATGVLPSPQKSTSLSKASPSBOEPL 1740  
1681 PRDKOKMOKELSELSTLSSLSRYFRGAPSSPATGVLPSPQKSTSLSKASPSBOEPL 1740

QY 1741 IQLVQAFVRHMOR 1753  
Db 1741 IQLVQAFVRHMOR 1753

RESULT 11  
US-10-719-385-16  
; Sequence 16, Application US/10719385  
; Publication No. US20040209284A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Toole et al.  
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
; FILE REFERENCE: 22058-582  
; CURRENT APPLICATION NUMBER: US/10/719,385  
; PRIOR FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: PCT/US03/37339  
; PRIOR FILING DATE: 2003-11-21  
; PRIOR FILING DATE: 2002-11-21  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 1753  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-719-385-16

Query Match 99.9%; Score 9001; DB 17; Length 1753;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1751; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFCRSSRELMTILGRSALRELSOIEAEINKMWRLLGSLYYKPPSPS 60  
Db 1 MIRSKITSVLSFCRSSRELMTILGRSALRELSOIEAEINKMWRLLGSLYYKPPSPS 60  
QY 61 SAEKRVANDVASPLKEIGIRISKEFLGDBEOSVOLLOCYLOEDYKTRDSVYTVLQDR 120  
Db 61 SAEKRVANDVASPLKEIGIRISKEFLGDBEOSVOLLOCYLOEDYKTRDSVYTVLQDR 120  
QY 121 QSQALILKADYEEERTCILACVHLILTYFQDERHPRYVEVADCVDKLEKIVSKYRQ 180  
Db 121 QSQALILKADYEEERTCILACVHLILTYFQDERHPRYVEVADCVDKLEKIVSKYRQ 180  
QY 181 FEEELYTEAPFWEETHNMLTERQVRSWRFVQCLREOSMLAEIFLYAYAFEMAPSDLVYT 240  
Db 181 FEEELYTEAPFWEETHNMLTERQVRSWRFVQCLREOSMLAEIFLYAYAFEMAPSDLVYT 240  
QY 241 KMFKEQGFSSROTRNLVDETNDPFVDRIGYFSALILVEGMDIESIAHKCALDRRELQF 300  
Db 241 KMFKEQGFSSROTRNLVDETNDPFVDRIGYFSALILVEGMDIESIAHKCALDRRELQF 300  
QY 301 AODGLICQMDICMLTFGDI PHHAPVILAMALLRHTLNFEETSVAVRKIGTRAIQUNVQ 360  
Db 301 AODGLICQMDICMLTFGDI PHHAPVILAMALLRHTLNFEETSVAVRKIGTRAIQUNVQ 360  
QY 361 YLTRLIOASAGNDCTTSTACMCYVGLSPVLTSLIELHTLNGOODIDTACEVLADESL 420  
Db 361 YLTRLIOASAGNDCTTSTACMCYVGLSPVLTSLIELHTLNGOODIDTACEVLADESL 420  
QY 421 PELFWGTEPTSGILGILDSVCMFPHLLSPLOQLRALVSGSKTAKKYSPFDKXSPYNE 480  
Db 421 PELFWGTEPTSGILGILDSVCMFPHLLSPLOQLRALVSGSKTAKKYSPFDKXSPYNE 480  
QY 481 LYKHKPHDVISHEGDTLWRQTPKLLYPLGGQTNLRIPQTVGYMLDBRAVLWMBEYSY 540  
Db 481 LYKHKPHDVISHEGDTLWRQTPKLLYPLGGQTNLRIPQTVGYMLDBRAVLWMBEYSY 540  
QY 541 SSWTLFTCEIEMLHVSTADVIQHCORVYKPIIDLVAHVISTDLSIADCLPITSRIYML 600  
Db 541 SSWTLFTCEIEMLHVSTADVIQHCORVYKPIIDLVAHVISTDLSIADCLPITSRIYML 600  
QY 601 LORLTVISPPVDVIAACVNCCLTVLAARNPAKWTDLRHGTGFLPVVAHPVSSLSQMSIAE 660

```

601 LQRLITVISPVDVIAQVNCILTVLAARNPAKWTDLRHGFLFVNAHPVSSLQMSIAE 660
661 GMAAGYGNLMLNSEQPOGEYGVTTAFRLITTVKQGLSTQGLVPCVMFLKEMLP 720
661 GMAAGYGNLMLNSEQPOGEYGVTTAFRLITTVKQGLSTQGLVPCVMFLKEMLP 720
721 SYHKWRVNSHGVROIGLITELJHAILINCHENDHSHSPSLOPCTCSLATTEAGOT 780
721 SYHKWRVNSHGVROIGLITELJHAILINCHENDHSHSPSLOPCTCSLATTEAGOT 780
781 VININGIGVDTIDVWMAAOPRSDAEGQOGLIKTVKLAFTYNNVIRLKPSNVSP 840
781 VININGIGVDTIDVWMAAOPRSDAEGQOGLIKTVKLAFTYNNVIRLKPSNVSP 840
841 LEQALSHGAGNNLIVAKYIYHKDPAIPRLAIQILKRLATVAPKSVACIGNDAAA 900
841 LEQALSHGAGNNLIVAKYIYHKDPAIPRLAIQILKRLATVAPKSVACIGNDAAA 900
901 IRDAFLRLQSKIEDMRKIMLEPLTVAVETOPGLIELFNLVKGSDGSKPSIGMW 960
901 IRDAFLRLQSKIEDMRKIMLEPLTVAVETOPGLIELFNLVKGSDGSKPSIGMW 960
961 SCLHAVIELIDSOQODRYWCPPLHRAAIAELHALMODRDSAMLVLRTEKFEWNTLSP 1020
961 SCLHAVIELIDSOQODRYWCPPLHRAAIAELHALMODRDSAMLVLRTEKFEWNTLSP 1020
1021 LFGTLPSPSETSEPSILETCALIMKIICLETIYVVVKGSLDQSLKDTLKFSIERPAWMS 1080
1021 LFGTLPSPSETSEPSILETCALIMKIICLETIYVVVKGSLDQSLKDTLKFSIERPAWMS 1080
1081 GYVSVLAVHVAETEGSSCTSLLEYOMLVSAWRMLIATTHADIMHLTDSVVRQOLFDV 1140
1081 GYVSVLAVHVAETEGSSCTSLLEYOMLVSAWRMLIATTHADIMHLTDSVVRQOLFDV 1140
1141 LDGKALLLVASVNCILGSMKCTLLIILRQWRELSGYDEILGPTLEILEGVLQADQ 1200
1141 LDGKALLLVASVNCILGSMKCTLLIILRQWRELSGYDEILGPTLEILEGVLQADQ 1200
1201 QLMKRYAKVSAFTTVLQMKEMKVSQDIPOYSQVLVNCETLQEBVIALPQTRHSLALG 1260
1201 QLMKRYAKVSAFTTVLQMKEMKVSQDIPOYSQVLVNCETLQEBVIALPQTRHSLALG 1260
1261 SATEDKSMETDCCSRSHRDQDGVCVLGLHAKELCEVEDSDSWLQVTRRLPIPLTL 1320
1261 SATEDKSMETDCCSRSHRDQDGVCVLGLHAKELCEVEDSDSWLQVTRRLPIPLTL 1320
1321 LTTLEVSILRMKONHFEATLHLITTLARTQOGATAVAGAGITOSICLPLSVYQLSTNG 1380
1321 LTTLEVSILRMKONHFEATLHLITTLARTQOGATAVAGAGITOSICLPLSVYQLSTNG 1380
1381 TAQTPSASRSKSLDAPSPMGVYRLSMSLMEQLLKTLYNPLPEALDFVGVHQRTELQCLNA 1440
1381 TAQTPSASRSKSLDAPSPMGVYRLSMSLMEQLLKTLYNPLPEALDFVGVHQRTELQCLNA 1440
1441 VRTVOSLACEEAHTGFTILOSNMKEKMHPLPOLMRDIOVNLGVYCOACTSLHSRK 1500
1441 VRTVOSLACEEAHTGFTILOSNMKEKMHPLPOLMRDIOVNLGVYCOACTSLHSRK 1500
1501 MLQHYLONKNGDGLPSAFAQROVOPPSAASAPSSSKOPADTEASEQOALHTVQYGLK 1560
1501 MLQHYLONKNGDGLPSAFAQROVOPPSAASAPSSSKOPADTEASEQOALHTVQYGLK 1560
1561 ILSTKTLAALRHFTDVCOILIDOSIDLAEYNPLPALSTTPTPPESEVAPSGFTLLATVNV 1620
1561 ILSTKTLAALRHFTDVCOILIDOSIDLAEYNPLPALSTTPTPPESEVAPSGFTLLATVNV 1620
1621 ALNMLGELDKKKEPLTOAVGSTQAGERTLKSLMFMENCFFYLLISQARRYLBDPVAH 1680
1621 ALNMLGELDKKKEPLTOAVGSTQAGERTLKSLMFMENCFFYLLISQARRYLBDPVAH 1680
1681 PRDKRMKQELSELSTLSSLSRYFRGAPSSPATVLPSPQKSTSLSKASPESSOEPL 1740
1681 PRDKRMKQELSELSTLSSLSRYFRGAPSSPATVLPSPQKSTSLSKASPESSOEPL 1740

```

```

Db 1681 PRDKRMKQELSELSTLSSLSRYFRGAPSSPATVLPSPQKSTSLSKASPESSOEPL 1740
Qy 1741 IOLVQAFVRHMQR 1753
Db 1741 IOLVQAFVRHMQR 1753

RESULT 12
US-10-719-385-7
; Sequence 7, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; PILE REFERENCE: 22058-582
; CURRENT FILING DATE: US/10/719,385
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-7

Query Match 99.9%; Score 8999; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1752; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MIRSKITTSVLSFCRSSRELTITLLGRSALRELSQIBALELKNWRRLLEGISTYKPPSPS 60
Db 1 MIRSKITTSVLSFCRSSRELTITLLGRSALRELSQIBALELKNWRRLLEGISTYKPPSPS 60
Qy 61 SAEKVRANKQVNASPLKELGLRISKFLGLDEBSVOLQCYLOEYRGTBDSKVTVLODER 120
Db 61 SAEKVRANKQVNASPLKELGLRISKFLGLDEBSVOLQCYLOEYRGTBDSKVTVLODER 120
Qy 121 QSOALILKADYYEERTCLIRCVLHLITTYFODERHRYRVEYADCVKLEKELVSKYRQ 180
Db 121 QSOALILKADYYEERTCLIRCVLHLITTYFODERHRYRVEYADCVKLEKELVSKYRQ 180
Qy 181 FEEELYKTEAPTWETHGNLMTERQVSRMFVQCLREQSMLEIIFLYAYFEMAPSDLLVLT 240
Db 181 FEEELYKTEAPTWETHGNLMTERQVSRMFVQCLREQSMLEIIFLYAYFEMAPSDLLVLT 240
Qy 241 KMFKEQGRSGQTRRHVDETMDFVDRIGYFSALLIVEGWDISLHKCALDDRREIHOQ 300
Db 241 KMFKEQGRSGQTRRHVDETMDFVDRIGYFSALLIVEGWDISLHKCALDDRREIHOQ 300
Qy 301 AODGLICQDMDCMLTFEGDIPHNAPVLLAWALLBHTLANPEBTSVVERKIGGTALQLVNFQ 360
Db 301 AODGLICQDMDCMLTFEGDIPHNAPVLLAWALLBHTLANPEBTSVVERKIGGTALQLVNFQ 360
Qy 361 YLTRILQSLASGGNDCTTSTACVCVYGLSFVLTSLBHTLGNODIITDACEVLADBSL 420
Db 361 YLTRILQSLASGGNDCTTSTACVCVYGLSFVLTSLBHTLGNODIITDACEVLADBSL 420
Qy 421 PELFMGTBPTSGGLIIDSVCGMFPLSLPLQLRLALVSGKSTAKKYVSFLDQMSFYNE 480
Db 421 PELFMGTBPTSGGLIIDSVCGMFPLSLPLQLRLALVSGKSTAKKYVSFLDQMSFYNE 480
Qy 481 LYKHKPHDIVSHEDGTLMRQTPRLLYPLGQOTMARIPOGTGQVMDLDRAYLVWRWESY 540
Db 481 LYKHKPHDIVSHEDGTLMRQTPRLLYPLGQOTMARIPOGTGQVMDLDRAYLVWRWESY 540
Qy 541 SSWTLFCEIEMLLHVVSTADVIOHCORVKEIIDLVRKVSTDSIADCLLPITSRIYML 600
Db 541 SSWTLFCEIEMLLHVVSTADVIOHCORVKEIIDLVRKVSTDSIADCLLPITSRIYML 600

```

```

QY 601 LORLTTVISPVDVIAVCNCLTVLAARNPAKWTDLRHTGFLPFAHVPASLSQMIAS 660
DB 601 LORLTTVISPVDVIAVCNCLTVLAARNPAKWTDLRHTGFLPFAHVPASLSQMIAS 660
QY 661 GNNAGGYGNLNLNSQPOGEVGTIAFLRLITTLVKGOLGSTOSGVLPCVMEVILKEMLP 720
DB 661 GNNAGGYGNLNLNSQPOGEVGTIAFLRLITTLVKGOLGSTOSGVLPCVMEVILKEMLP 720
QY 721 SYHKRNVSHGVREOIGCLIELHAHNLCHETLHSHSTPSLOFLICISLAATYEAQT 780
DB 721 SYHKRNVSHGVREOIGCLIELHAHNLCHETLHSHSTPSLOFLICISLAATYEAQT 780
QY 781 VININGIGVDTIDWMAAOPRSDGAEQGOQGLIKTVKLAFTVNTNVIIRLKPSSNVSP 840
DB 781 VININGIGVDTIDWMAAOPRSDGAEQGOQGLIKTVKLAFTVNTNVIIRLKPSSNVSP 840
QY 841 LEQALSQGHAGNNILAVLAKYIKHDPALPRLAIOCLKRLATVAPWSVYACIGNDAAA 900
DB 841 LEQALSQGHAGNNILAVLAKYIKHDPALPRLAIOCLKRLATVAPWSVYACIGNDAAA 900
QY 901 IRDAFLTLQSKIEDMRIKMILEFLVAVETOPGLIELFLNLEVKOSDGSKEPSLGMW 960
DB 901 IRDAFLTLQSKIEDMRIKMILEFLVAVETOPGLIELFLNLEVKOSDGSKEPSLGMW 960
QY 961 SCLHNVIELIDSQODRYWCPEPLHRAAIAFLHLMODRDSAMLVLRKPKFMENTLSP 1020
DB 961 SCLHNVIELIDSQODRYWCPEPLHRAAIAFLHLMODRDSAMLVLRKPKFMENTLSP 1020
QY 1021 LFGTLPSPSESEBSIETCALIMKITLCTEYVVYKSLDLSLDTLKKSEIERFAYWS 1080
DB 1021 LFGTLPSPSESEBSIETCALIMKITLCTEYVVYKSLDLSLDTLKKSEIERFAYWS 1080
QY 1081 GYVSLAVHVAETEGSSCTSLLEYOMLVSAARMMLIIATTHADIMHLDTSVVRQLFLDV 1140
DB 1081 GYVSLAVHVAETEGSSCTSLLEYOMLVSAARMMLIIATTHADIMHLDTSVVRQLFLDV 1140
QY 1141 LDGKALLVPAVSNCLRLSGMKCTLLILRLKMRKELSGVDELGLPTEILBEGVADQ 1200
DB 1141 LDGKALLVPAVSNCLRLSGMKCTLLILRLKMRKELSGVDELGLPTEILBEGVADQ 1200
QY 1201 QLMKTRAKVSAFTVLQMKEMKVSIPQYSOLVAVNCELTQHEVIALPQTRHSIALG 1260
DB 1201 QLMKTRAKVSAFTVLQMKEMKVSIPQYSOLVAVNCELTQHEVIALPQTRHSIALG 1260
QY 1261 SATEDKDSMETDSCSRSHRDQDQVCYLGLHAKELCEVEDSDSWLOVTRRLPIPLT 1320
DB 1261 SATEDKDSMETDSCSRSHRDQDQVCYLGLHAKELCEVEDSDSWLOVTRRLPIPLT 1320
QY 1321 LTTLEVLBMKQNHFEATLHLLTLARTOOGATVAGAGITOSTCLPLSTYQSTNG 1380
DB 1321 LTTLEVLBMKQNHFEATLHLLTLARTOOGATVAGAGITOSTCLPLSTYQSTNG 1380
QY 1381 TAQTPSASRSKSLDAPSPGVYRLSMSLMBOLIKTLRNLPEALDFGVGHOERTLOCLNA 1440
DB 1381 TAQTPSASRSKSLDAPSPGVYRLSMSLMBOLIKTLRNLPEALDFGVGHOERTLOCLNA 1440
QY 1441 VRTVQSLACLEADHTVGFILQLSNFKEMWHFILPOLMRDIOVNLGYLCOACTSLHSRK 1500
DB 1441 VRTVQSLACLEADHTVGFILQLSNFKEMWHFILPOLMRDIOVNLGYLCOACTSLHSRK 1500
QY 1501 MLQHYLONKQKGDGPSAVQVORPPPSAASAPSSSQPADTEASQOALHTVOYGLK 1560
DB 1501 MLQHYLONKQKGDGPSAVQVORPPPSAASAPSSSQPADTEASQOALHTVOYGLK 1560
QY 1561 ILSKTLAALRHFTPDVCOILLDOSLDLAENFLFALSFTTPTDSEVAPSFGLTALVNY 1620
DB 1561 ILSKTLAALRHFTPDVCOILLDOSLDLAENFLFALSFTTPTDSEVAPSFGLTALVNY 1620
QY 1621 ALNMLGELDKKKEBELTQAVGLSTQABESTRTLSLMTMENCYLLISQAMRYLRPAAVA 1680
DB 1621 ALNMLGELDKKKEBELTQAVGLSTQABESTRTLSLMTMENCYLLISQAMRYLRPAAVA 1680
QY 1681 PRDKORRKOBSLSELSLSSLSRYFRKGA PSSPATVLPSPQKSTSLSKAPSESOEPL 1740

```

```

DB 1681 PRDKORRKOBSLSELSLSSLSRYFRKGA PSSPATVLPSPQKSTSLSKAPSESOEPL 1740
QY 1741 IOLVOAFVRMOR 1753
DB 1741 IOLVOAFVRMOR 1753

RESULT 13
US-10-719-385-12
; Sequence 12, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-12

Query Match          99.9%; Score 8999; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1752; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MIRSKITTSVLSFCRSGRELMWTLTGRSALRELSQIEALNKRRLLEGSLYKPPSPS 60
DB 1 MIRSKITTSVLSFCRSGRELMWTLTGRSALRELSQIEALNKRRLLEGSLYKPPSPS 60
QY 61 SAEKVANDVAPSLKELGLRISKFLGDEBOSVOLQCYLOEDYVGTDSVTVQODR 120
DB 61 SAEKVANDVAPSLKELGLRISKFLGDEBOSVOLQCYLOEDYVGTDSVTVQODR 120
QY 121 OSQALILKADYYEERTCILRCVHLITTYFODERHPYREYVADCVKLEKELVSKYRQ 180
DB 121 OSQALILKADYYEERTCILRCVHLITTYFODERHPYREYVADCVKLEKELVSKYRQ 180
QY 181 FEELYKTEAPYETHGNLWTERQVSRWFQCIREQSMLEITFLYYAYFEMAPSDLLVIT 240
DB 181 FEELYKTEAPYETHGNLWTERQVSRWFQCIREQSMLEITFLYYAYFEMAPSDLLVIT 240
QY 241 KMEKEGFSROTRNLVDETMDPFVDRIGYFSALILVGMDESLHRCALDRRELHOF 300
DB 241 KMEKEGFSROTRNLVDETMDPFVDRIGYFSALILVGMDESLHRCALDRRELHOF 300
QY 301 AODGLICQDMDCIMLTFGDI PHRAPVLLAMALLRHTLNEETS SVVRKIGTAIQLNVQ 360
DB 301 AODGLICQDMDCIMLTFGDI PHRAPVLLAMALLRHTLNEETS SVVRKIGTAIQLNVQ 360
QY 361 YLTRLLQSLASGANDCTTSTACMCYVGLSFLVTSLEHTLGNODIITPTACEVLADEPL 420
DB 361 YLTRLLQSLASGANDCTTSTACMCYVGLSFLVTSLEHTLGNODIITPTACEVLADEPL 420
QY 421 PELFWGTEPTSGIILDSVCGMFPHLSPFLQILRALVSGSKTKKYYSPFDKMSFYNE 480
DB 421 PELFWGTEPTSGIILDSVCGMFPHLSPFLQILRALVSGSKTKKYYSPFDKMSFYNE 480
QY 481 LYKHPHVDVISHEDGTLMWRQTPKLLYPLGQGTNLRIPQTVGQVMLDRAVLVMEVSY 540
DB 481 LYKHPHVDVISHEDGTLMWRQTPKLLYPLGQGTNLRIPQTVGQVMLDRAVLVMEVSY 540
QY 541 SSWTLFTGCEIMLLVSVSTADYIQHCQVKKPIIDLVHVKVISTDLSIADCLPITRSIYML 600
DB 541 SSWTLFTGCEIMLLVSVSTADYIQHCQVKKPIIDLVHVKVISTDLSIADCLPITRSIYML 600

```

```

QY 601 LQRLITTVISPPVDYIASCNCLITVLAARNPAKWTDLHGTFLPFAVPSVLSOMISAB 660
Db 601 LQRLITTVISPPVDYIASCNCLITVLAARNPAKWTDLHGTFLPFAVPSVLSOMISAB 660
QY 661 GMAAGYGNILMNSBOPQGEYGVITIAFLRLITTVKQGLSTQSGVLPCVMFVLEKMLP 720
Db 661 GMAAGYGNILMNSBOPQGEYGVITIAFLRLITTVKQGLSTQSGVLPCVMFVLEKMLP 720
QY 721 SYHKWRINSHGVRQIGCLIELIHALIINLCHETDHSHTPSIQFLCTIGSLATTEAGOT 780
Db 721 SYHKWRINSHGVRQIGCLIELIHALIINLCHETDHSHTPSIQFLCTIGSLATTEAGOT 780
QY 781 VINIMGIGVDITIDWMAAOPRSDGAEQGGQGLIKITKLAFSVTNNVIRLKPSSNVSP 840
Db 781 VINIMGIGVDITIDWMAAOPRSDGAEQGGQGLIKITKLAFSVTNNVIRLKPSSNVSP 840
QY 841 LEQALSOHGAHGNLILAVLAKYIYHKDPAFLRLIQLKSLATVAPMSVYACIGNDAAA 900
Db 841 LEQALSOHGAHGNLILAVLAKYIYHKDPAFLRLIQLKSLATVAPMSVYACIGNDAAA 900
QY 901 IRDAFLTRLQSKIEDMRIKWILEFLITVAVETOPGLIELFLNLEVGQSDGSKESLTMM 960
Db 901 IRDAFLTRLQSKIEDMRIKWILEFLITVAVETOPGLIELFLNLEVGQSDGSKESLTMM 960
QY 961 SCLHVLIELIDSOODRWCPPLHRAIAFLHLMODRDSAMLVLRTPKPMENLTSP 1020
Db 961 SCLHVLIELIDSOODRWCPPLHRAIAFLHLMODRDSAMLVLRTPKPMENLTSP 1020
QY 1021 LFGTLSPSESTSEPSIETCALIMKICLEIYVVYVKSGLDQSLKDTLKFSIEKRFAYS 1080
Db 1021 LFGTLSPSESTSEPSIETCALIMKICLEIYVVYVKSGLDQSLKDTLKFSIEKRFAYS 1080
QY 1081 GYVVSILAVHAEFTGSSCTSLLEYOMLVASAMRMLIITTHADIMHLDTSVYRROLFDV 1140
Db 1081 GYVVSILAVHAEFTGSSCTSLLEYOMLVASAMRMLIITTHADIMHLDTSVYRROLFDV 1140
QY 1141 LDGFKALLVPAVAVNCRLGSMKCTLLILRLKRMKRELSYDEILGPTLEILEGVLQDO 1200
Db 1141 LDGFKALLVPAVAVNCRLGSMKCTLLILRLKRMKRELSYDEILGPTLEILEGVLQDO 1200
QY 1201 QLMEXTKAVFSAPITVLQMKEMKVSIPQYSQLVAVNCETLQEBEVALPQTRHSIALG 1260
Db 1201 QLMEXTKAVFSAPITVLQMKEMKVSIPQYSQLVAVNCETLQEBEVALPQTRHSIALG 1260
QY 1261 SATEDKDSMETDDCSRSRHRDQRCVGLGHLAKELCEVDEDDGSMVQVTRRLPIPLTL 1320
Db 1261 SATEDKDSMETDDCSRSRHRDQRCVGLGHLAKELCEVDEDDGSMVQVTRRLPIPLTL 1320
QY 1321 LTTLEVSILRMKONLHFTBATLHLITLRLARQOGAFAVAGAGITOSICLPLSVYQVLTNG 1380
Db 1321 LTTLEVSILRMKONLHFTBATLHLITLRLARQOGAFAVAGAGITOSICLPLSVYQVLTNG 1380
QY 1381 TAQTPSASRKSILDAFSPGVRRLSWSLMEQLKTLRYNFLPEALDFVGHQERTLOCLNA 1440
Db 1381 TAQTPSASRKSILDAFSPGVRRLSWSLMEQLKTLRYNFLPEALDFVGHQERTLOCLNA 1440
QY 1441 VRTVQSLACLEADHTVGFIIQLSNFKMKWFHFLPOLMRDIOVNLGICYQAQCTSLHSRK 1500
Db 1441 VRTVQSLACLEADHTVGFIIQLSNFKMKWFHFLPOLMRDIOVNLGICYQAQCTSLHSRK 1500
QY 1501 MLQHYLONKNGDGLPSAFAORVORPPSASAAPSSSKOPADPTESSEQALHTVOYGLK 1560
Db 1501 MLQHYLONKNGDGLPSAFAORVORPPSASAAPSSSKOPADPTESSEQALHTVOYGLK 1560
QY 1561 ILSTKLALRHFTPDVCOILLDQSDILAENYFLFALSTPTPFSSEVAPSGFTLLATVNV 1620
Db 1561 ILSTKLALRHFTPDVCOILLDQSDILAENYFLFALSTPTPFSSEVAPSGFTLLATVNV 1620
QY 1621 ALANMIGEIDKKKEPTQAVGLSTQAEGRITLKSILMFTMENCYLLISQAMRYLRDPVH 1680
Db 1621 ALANMIGEIDKKKEPTQAVGLSTQAEGRITLKSILMFTMENCYLLISQAMRYLRDPVH 1680

```

```

QY 1681 PRDKORMKOBLSSELSTLSSLSRYFRGAPSSPATGVLSPQSKSTLSKASPESEQPL 1740
Db 1681 PRDKORMKOBLSSELSTLSSLSRYFRGAPSSPATGVLSPQSKSTLSKASPESEQPL 1740
QY 1741 IQLVOAFVRHMOR 1753
Db 1741 IQLVOAFVRHMOR 1753

RESULT 14
US-10-719-385-18
; Sequence 18, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1753
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-719-385-18

Query Match 99.9%; Score 8998; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1751; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFCRSSREIMWTIILGRSALRELSQIEAEINKMRRLLEGISYKPPSPS 60
Db 1 MIRSKITSVLSFCRSSREIMWTIILGRSALRELSQIEAEINKMRRLLEGISYKPPSPS 60
QY 61 SAEVYKANKVVASPLKELGLRISKFLGIDEQSVQLQCYQOEYRGTRDSVKTVLODER 120
Db 61 SAEVYKANKVVASPLKELGLRISKFLGIDEQSVQLQCYQOEYRGTRDSVKTVLODER 120
QY 121 OSQALLIKIADYYEERTCIRCVLHLLTYQODRRHPFRVAYACVDKLEELVSKTRQO 180
Db 121 OSQALLIKIADYYEERTCIRCVLHLLTYQODRRHPFRVAYACVDKLEELVSKTRQO 180
QY 181 FEEELYKTEAPTWETHGNLMTEROVSRYMFVQCLREOSMLBEIIPYAYAFENAPSDLVLT 240
Db 181 FEEELYKTEAPTWETHGNLMTEROVSRYMFVQCLREOSMLBEIIPYAYAFENAPSDLVLT 240
QY 241 KMFEKQGFSGQTRNRLVDETMDFVDRIGFSALLIVEGMDISLHKCALDDRRELHQF 300
Db 241 KMFEKQGFSGQTRNRLVDETMDFVDRIGFSALLIVEGMDISLHKCALDDRRELHQF 300
QY 301 AODGLICQDMDCMLTFGDIPIHHA PVLAMALBHTLNPESTSVRKIGGTALQNAVFO 360
Db 301 AODGLICQDMDCMLTFGDIPIHHA PVLAMALBHTLNPESTSVRKIGGTALQNAVFO 360
QY 361 YLTRLLQSLASGNDCTTSTA CMKVYGLSFTLSLEHTLGNQODIIDTACEVLADBSL 420
Db 361 YLTRLLQSLASGNDCTTSTA CMKVYGLSFTLSLEHTLGNQODIIDTACEVLADBSL 420
QY 421 PELFWGTEPTSGLGIIIDSVCGMFPPLISPLQLRLALVSGKSTAKKYVSLDKKSPFNE 480
Db 421 PELFWGTEPTSGLGIIIDSVCGMFPPLISPLQLRLALVSGKSTAKKYVSLDKKSPFNE 480
QY 481 LYKXKPHDIVSHEDGTLMBRQTPKLVPLVGGQTNLRIPQGTVGQVMDDBRAYLVKWEYSY 540
Db 481 LYKXKPHDIVSHEDGTLMBRQTPKLVPLVGGQTNLRIPQGTVGQVMDDBRAYLVKWEYSY 540
QY 541 SSWTLFTCEIEMLLHVSTADVIOHCORVKPDI DLVHKVISTDLSIADCLLPITSRIYML 600

```



Db 541 SSWTLFTCEIEMLHVSTADVIQHQRVKPIIDLHVKVIISTDLSIADCLPITSRIYML 600  
 Qy 601 LQRLTIVISPPVDIVASCNCLTIVLAARNPAKWTDLRHTEGFLPVAHVPSSLQMSISAE 660  
 Db 601 LQRLTIVISPPVDIVASCNCLTIVLAARNPAKWTDLRHTEGFLPVAHVPSSLQMSISAE 660  
 Qy 661 GMAAGYGNLMAANSPOGOEYGTIAFLRLITLVKGOGLSGTOSQGLPCWCFVTKMLP 720  
 Db 661 GMAAGYGNLMAANSPOGOEYGTIAFLRLITLVKGOGLSGTOSQGLPCWCFVTKMLP 720  
 Qy 721 SYHKRNVSHGVEQIGCLILELHAIINLCHETDLSHSHTPSLQFLCISLAATEAGOT 780  
 Db 721 SYHKRNVSHGVEQIGCLILELHAIINLCHETDLSHSHTPSLQFLCISLAATEAGOT 780  
 Qy 781 VINIMGIVDTIDMMAOPSPDGAEGGGGGLIKTKYKAPSVYNNIRIKPSPNVSP 840  
 Db 781 VINIMGIVDTIDMMAOPSPDGAEGGGGGLIKTKYKAPSVYNNIRIKPSPNVSP 840  
 Qy 841 LEQALSOHGAHGNMLIAVLAKYIYKHDPALPRLAIQCLKRLATVAPMSVYACLGNDAAA 900  
 Db 841 LEQALSOHGAHGNMLIAVLAKYIYKHDPALPRLAIQCLKRLATVAPMSVYACLGNDAAA 900  
 Qy 901 IRDAFLTLQSKIEEMRIKVMILEPITVAVTQELIFLNLVYKDGSDGSKESPISGMW 960  
 Db 901 IRDAFLTLQSKIEEMRIKVMILEPITVAVTQELIFLNLVYKDGSDGSKESPISGMW 960  
 Qy 961 SCLHVLIELIDSOODRYMCPPLHRAAIAFLHLMODRDSAMLVJTRKPEMINTSP 1020  
 Db 961 SCLHVLIELIDSOODRYMCPPLHRAAIAFLHLMODRDSAMLVJTRKPEMINTSP 1020  
 Qy 1021 LFGTISPPSETSEPSIETCALIMKIICLETIYVYKGSJDSQSLXDTLKKPSIEKRFAYMS 1080  
 Db 1021 LFGTISPPSETSEPSIETCALIMKIICLETIYVYKGSJDSQSLXDTLKKPSIEKRFAYMS 1080  
 Qy 1081 GYVSLAHVAHETGSSCTSLIETQMLVSARMLLIATTHADIMHLDTSVVRQQLFDV 1140  
 Db 1081 GYVSLAHVAHETGSSCTSLIETQMLVSARMLLIATTHADIMHLDTSVVRQQLFDV 1140  
 Qy 1141 LDGFKALLVPAVNCRLGSMKCTLLILIRKMKRELSGVDEILGPTEILIEGVLAQDQ 1200  
 Db 1141 LDGFKALLVPAVNCRLGSMKCTLLILIRKMKRELSGVDEILGPTEILIEGVLAQDQ 1200  
 Qy 1201 QLMKRTAKVSAITTVLQMKEMKVSIPQYSQVLINVCETLQEBVIALPDQTHSLALG 1260  
 Db 1201 QLMKRTAKVSAITTVLQMKEMKVSIPQYSQVLINVCETLQEBVIALPDQTHSLALG 1260  
 Qy 1261 SATEDXOSMETDSCSRHRDQDQVCVGLHMLAKECEVEDGDSMLQVTRRLPIPLTL 1320  
 Db 1261 SATEDXOSMETDSCSRHRDQDQVCVGLHMLAKECEVEDGDSMLQVTRRLPIPLTL 1320  
 Qy 1321 LTTLEVSILRMKONLHFEATLHLLTLARTQOGATAVAGAITOSICLPILSYVQJSTNG 1380  
 Db 1321 LTTLEVSILRMKONLHFEATLHLLTLARTQOGATAVAGAITOSICLPILSYVQJSTNG 1380  
 Qy 1381 TAOPPSASRSKSLDAPSWRGVYRLSMSLMEQLKTLRNPFLPEALDPFGVHOERTLOCLNA 1440  
 Db 1381 TAOPPSASRSKSLDAPSWRGVYRLSMSLMEQLKTLRNPFLPEALDPFGVHOERTLOCLNA 1440  
 Qy 1441 VRTVQSLACLAEADHTGFILOLSNFMKEWHFHLPOLMRDIQVNLGILCOACTSLHLSRK 1500  
 Db 1441 VRTVQSLACLAEADHTGFILOLSNFMKEWHFHLPOLMRDIQVNLGILCOACTSLHLSRK 1500  
 Qy 1501 MLQHYLONKXGKDGIPSAVAQVORPPSAASAAPSSSTOPAADTEASBOQALHTVOYGLK 1560  
 Db 1501 MLQHYLONKXGKDGIPSAVAQVORPPSAASAAPSSSTOPAADTEASBOQALHTVOYGLK 1560  
 Qy 1561 ILKXTLALAHFPPDVCOIILLDGLDAEYNPLPALSFTTPTDSEVTAPEFGTILATVNY 1620  
 Db 1561 ILKXTLALAHFPPDVCOIILLDGLDAEYNPLPALSFTTPTDSEVTAPEFGTILATVNY 1620  
 Qy 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRITLKSILMTMENCFLYLSQAMRYLDPAYH 1680  
 Db 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRITLKSILMTMENCFLYLSQAMRYLDPAYH 1680

Qy 1681 PRDKORMKOBLSSELSTLSSLSRYFRGAPSSPATGVLPSPOCKSTLSKASPEQEP 1740  
 Db 1681 PRDKORMKOBLSSELSTLSSLSRYFRGAPSSPATGVLPSPOCKSTLSKASPEQEP 1740  
 Qy 1741 IOLVOAFVRHMOR 1753  
 Db 1741 IOLVOAFVRHMOR 1753  
 RESULT 15  
 US-10-719-385-9  
 ; Sequence 9, Application US/10719385  
 ; Publication No. US20040209284A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Toole et al.  
 ; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 ; FILE REFERENCE: 22058-582  
 ; CURRENT APPLICATION NUMBER: US/10/719,385  
 ; PRIOR APPLICATION NUMBER: PCT/US03/37339  
 ; PRIOR FILING DATE: 2003-11-21  
 ; PRIOR APPLICATION NUMBER: 60/428,094  
 ; PRIOR FILING DATE: 2002-11-21  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 9  
 ; LENGTH: 1753  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-719-385-9  
 Query Match 99.9%; Score 8997; DB 17; Length 1753;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MIRSKITSVLSFCRSRREIMTILGRSALRELSQIEAEINKMWRILGSLSYKPPSP 60  
 Db 1 MIRSKITSVLSFCRSRREIMTILGRSALRELSQIEAEINKMWRILGSLSYKPPSP 60  
 Qy 61 SAEKRVANDVASPLKELGIRISKFLGIDBEOVSQVLQCYLOEDYNGTDSVYTVLQDER 120  
 Db 61 SAEKRVANDVASPLKELGIRISKFLGIDBEOVSQVLQCYLOEDYNGTDSVYTVLQDER 120  
 Qy 121 QSOALILKXADYYEERTCILRCVHLITVPDERHPYVEYADCDKLEKELVSKYRQ 180  
 Db 121 QSOALILKXADYYEERTCILRCVHLITVPDERHPYVEYADCDKLEKELVSKYRQ 180  
 Qy 181 FEELYTEAPVWETHGNLMTERQVSRWFVQCUREQSMLEIIFLYAYEFEMAPSDILVLT 240  
 Db 181 FEELYTEAPVWETHGNLMTERQVSRWFVQCUREQSMLEIIFLYAYEFEMAPSDILVLT 240  
 Qy 241 KMKKEGFSRQTNRLVDETMDFVDRIGYSALILVEGMDIESLHKCALDDREILHOF 300  
 Db 241 KMKKEGFSRQTNRLVDETMDFVDRIGYSALILVEGMDIESLHKCALDDREILHOF 300  
 Qy 301 AODGLICQDMDCMLTFGDI PHHAPVILAMALIRHTINEESTSVYRKIGGTAICQANVO 360  
 Db 301 AODGLICQDMDCMLTFGDI PHHAPVILAMALIRHTINEESTSVYRKIGGTAICQANVO 360  
 Qy 361 YLTRLLQSLASGANDCTTSTACMCVYGLISFVLSLEHTLGNODIIDTACEVLADPSL 420  
 Db 361 YLTRLLQSLASGANDCTTSTACMCVYGLISFVLSLEHTLGNODIIDTACEVLADPSL 420  
 Qy 421 PELFWGTEPTSGILITLISVCCMFHLSPLILQLRALVSGSTAKVYSFLDKXSFYNE 480  
 Db 421 PELFWGTEPTSGILITLISVCCMFHLSPLILQLRALVSGSTAKVYSFLDKXSFYNE 480  
 Qy 481 LYKHKPHDIVISHEDEGLMRQPKLLYPAGQTNLRIPOGTGVQWMLDRAVLVMEYSY 540  
 Db 481 LYKHKPHDIVISHEDEGLMRQPKLLYPAGQTNLRIPOGTGVQWMLDRAVLVMEYSY 540  
 Qy 541 SSWTLFTCEIEMLHVSTADVIQHQRVKPIIDLHVKVIISTDLSIADCLPITSRIYML 600



Db 541 SSWLFLCEIEMLAHVSTADVIOHCQVRFIDLVKHXVISTDLSIDCLLPISRTYML 600  
Qy 601 LQRLITVISPVDVIAVCNCLTVLAARNPAKWTDLRHGFLPFVAHPVSSLSQMTSAB 660  
Db 601 LQRLITVISPVDVIAVCNCLTVLAARNPAKVTDLRHGFLPFVAHPVSSLSQMTSAB 660  
Qy 661 GMAAGGVNLLMSEOGEGVGTIAFLRLITTVKQGLSTOSOGIVPCMFWLKXMLP 720  
Db 661 GMAAGGVNLLMSEOGEGVGTIAFLRLITTVKQGLSTOSOGIVPCMFWLKXMLP 720  
Qy 721 SYHKMYNSHGVREIQGLILELITHAILNCHETDLSHSTPSIQFLICSLAYTEAGQT 780  
Db 721 SYHKMYNSHGVREIQGLILELITHAILNCHETDLSHSTPSIQFLICSLAYTEAGQT 780  
Qy 781 VINIMIGVDTIDMVAAPRSDGAEQOGOLIKTVKLAFSVTNNVIRLKPSSNVSP 840  
Db 781 VINIMIGVDTIDMVAAPRSDGAEQOGOLIKTVKLAFSVTNNVIRLKPSSNVSP 840  
Qy 841 LEQALSHQAHGNNLIVLAKYIYHKDPALPRLAIDLRKLAIVAPMSVYACLGNDAAA 900  
Db 841 LEQALSHQAHGNNLIVLAKYIYHKDPALPRLAIDLRKLAIVAPMSVYACLGNDAAA 900  
Qy 901 IRDAFLRLQSKIEDMRKIMVLEFLTVAVETQGLIELFNLBEVKDSDSKESFSLGMV 960  
Db 901 IRDAFLRLQSKIEDMRKIMVLEFLTVAVETQGLIELFNLBEVKDSDSKESFSLGMV 960  
Qy 961 SCHAVLELIDSOQDRYWCPLLHRAAIAFLAALMODRDSAMLVLRKPKFMENTSP 1020  
Db 961 SCHAVLELIDSOQDRYWCPLLHRAAIAFLAALMODRDSAMLVLRKPKFMENTSP 1020  
Qy 1021 LFGTLSPSESESESIETCLIMKITCLEIYVYVKSIDSLDOLTKKSEIERFVWS 1080  
Db 1021 LFGTLSPSESESESIETCLIMKITCLEIYVYVKSIDSLDOLTKKSEIERFVWS 1080  
Qy 1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSAMRMILLIATTTADIMHLDTSVVRQLFLDV 1140  
Db 1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSAMRMILLIATTTADIMHLDTSVVRQLFLDV 1140  
Qy 1141 LDGTKALLVPAVAVNCLRLSGMKCTLLILLRQMKREIGSVDEILGPULTEILEVQADQ 1200  
Db 1141 LDGTKALLVPAVAVNCLRLSGMKCTLLILLRQMKREIGSVDEILGPULTEILEVQADQ 1200  
Qy 1201 QLMKTKAKYFSAITVLOMKEMKVSIDIPOYSQVAVNCEFLQREVALPOTGHSIALG 1260  
Db 1201 QLMKTKAKYFSAITVLOMKEMKVSIDIPOYSQVAVNCEFLQREVALPOTGHSIALG 1260  
Qy 1261 SATEDKSMETDSCSRSHRDQDGVCVLGLHLAKELCEVEDDGSMLQVTRRLPILPTL 1320  
Db 1261 SATEDKSMETDSCSRSHRDQDGVCVLGLHLAKELCEVEDDGSMLQVTRRLPILPTL 1320  
Qy 1321 LTTLEVLIRMKONLHFTETATLHLLTLARTOGATAVAGAGITQSI CILPLISVYQLSTNG 1380  
Db 1321 LTTLEVLIRMKONLHFTETATLHLLTLARTOGATAVAGAGITQSI CILPLISVYQLSTNG 1380  
Qy 1381 TAQCPSSSRSLDAPSPWGVYRLSMSLEQLLKTARVNFLEALDFGVHQBERTLOCLANA 1440  
Db 1381 TAQCPSSSRSLDAPSPWGVYRLSMSLEQLLKTARVNFLEALDFGVHQBERTLOCLANA 1440  
Qy 1441 VRTVQSLACEAHTVGFILQLSNFMKEWHFHLPOLMRDIOVNLGYCOACTSLHSRK 1500  
Db 1441 VRTVQSLACEAHTVGFILQLSNFMKEWHFHLPOLMRDIOVNLGYCOACTSLHSRK 1500  
Qy 1501 MLQHYLONKNGDGLPSAVALQVQRPSPSAAPSSSKQPADTEASEQOALHTVQYGLK 1560  
Db 1501 MLQHYLONKNGDGLPSAVALQVQRPSPSAAPSSSKQPADTEASEQOALHTVQYGLK 1560  
Qy 1561 ILSTTLAALHRTFDVQOILLDOSLDLAEVNFALSTTTTPPSEVAPSGCTILLATNV 1620  
Db 1561 ILSTTLAALHRTFDVQOILLDOSLDLAEVNFALSTTTTPPSEVAPSGCTILLATNV 1620  
Qy 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRTLKSLMFTMENC FYLLISQAMRYLRDPVH 1680

Db 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRTLKSLMFTMENC FYLLISQAMRYLRDPVH 1680  
Qy 1681 PRDQKRNKQELSSSTLSTLSLSKTYFRRCAPSSSRTATVLSPOCKSTLSKASPEQOEP 1740  
Db 1681 PRDQKRNKQELSSSTLSTLSLSKTYFRRCAPSSSRTATVLSPOCKSTLSKASPEQOEP 1740  
Qy 1741 IOLVQAFVRHMOR 1753  
Db 1741 IOLVQAFVRHMOR 1753  
  
RESULT 16  
US-10-719-385-17  
; Sequence 17, Application US/10719385  
; Publication No. US20040209284A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Toole et al.  
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
; FILE REFERENCE: 22058-582  
; CURRENT APPLICATION NUMBER: US/10/719,385  
; PRIOR FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: PCT/US03/37339  
; PRIOR FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: 60/428,094  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 1753  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-719-385-17  
  
Query Match 99.8%; Score 8993; DB 17; Length 1753;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1751; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 MIRSKITSVLSPCRSSRELWTILGRSALRELSQIEAEINKHRRLLBGLSYKPPSPS 60  
Db 1 MIRSKITSVLSPCRSSRELWTILGRSALRELSQIEAEINKHRRLLBGLSYKPPSPS 60  
Qy 61 SAEKYKANKQVAPSLKELGLIRISKFGLDEBQSVQLQCYLOEDYRGTBDSVKTVLDER 120  
Db 61 SAEKYKANKQVAPSLKELGLIRISKFGLDEBQSVQLQCYLOEDYRGTBDSVKTVLDER 120  
Qy 121 QSQALLIKIADYEEERTCILRCVLAHLITYFODERHRYRVEYADCVKLEKELVSKTRQ 180  
Db 121 QSQALLIKIADYEEERTCILRCVLAHLITYFODERHRYRVEYADCVKLEKELVSKTRQ 180  
Qy 181 FEEIYKTEAPWETHGNLMTEROYSRWFOCLARQSMLEIIFLYVAYFEMAPSDLLVLT 240  
Db 181 FEEIYKTEAPWETHGNLMTEROYSRWFOCLARQSMLEIIFLYVAYFEMAPSDLLVLT 240  
Qy 241 KMFKEQFGSROTNRHLVDETMDPFVDRIGYFSALLIVEGMDISLHKCALDDRRRLHQF 300  
Db 241 KMFKEQFGSROTNRHLVDETMDPFVDRIGYFSALLIVEGMDISLHKCALDDRRRLHQF 300  
Qy 301 AODGLIQDMDCLMLTFGDIPEHAPVLLAVALLHRTLNPEETSVAVKIGGTAIQLNVFQ 360  
Db 301 AODGLIQDMDCLMLTFGDIPEHAPVLLAVALLHRTLNPEETSVAVKIGGTAIQLNVFQ 360  
Qy 361 YLTPLLQSLASGNDCTTSTACMKCVGLSLSVLTSLEHLTNGNODIIPTRACVLAAPSL 420  
Db 361 YLTPLLQSLASGNDCTTSTACMKCVGLSLSVLTSLEHLTNGNODIIPTRACVLAAPSL 420  
Qy 421 PELFWGTEPTSGGLIILDSVCGMFPHLSPILQLLRALVSGKSTAKKYVSFLDMSFYNE 480  
Db 421 PELFWGTEPTSGGLIILDSVCGMFPHLSPILQLLRALVSGKSTAKKYVSFLDMSFYNE 480  
Qy 481 LYKXKPHDIVISHEDGTLMRQTPLLYPLGQTNLRIPOGTGQVMDLDRAYLVREYSY 540  
Db 481 LYKXKPHDIVISHEDGTLMRQTPLLYPLGQTNLRIPOGTGQVMDLDRAYLVREYSY 540

```

QY 541 SSWTLFCEIEMLTAVSTADYIOHCORVKPIIDLVHKVISTDLSIADCLIPITSRIYML 600
DB 541 SSWTLFCEIEMLTAVSTADYIOHCORVKPIIDLVHKVISTDLSIADCLIPITSRIYML 600
QY 601 LQRLTTVISPVDVIASCVNCLTVLAARNPAKWTDLHRTGFLPFAHVPBSLSQMTSAE 660
DB 601 LQRLTTVISPVDVIASCVNCLTVLAARNPAKWTDLHRTGFLPFAHVPBSLSQMTSAE 660
QY 661 GMMNGGVNLTMSNPOGEYGVTAFLRTLTLYKGLSTGSGVLPCVMFYLKMLP 720
DB 661 GMMNGGVNLTMSNPOGEYGVTAFLRTLTLYKGLSTGSGVLPCVMFYLKMLP 720
QY 721 SYHKRYNSHGVREGICLTLELHAINTLCHETDHSHTPSLOFLCTISLAYTEAGOT 780
DB 721 SYHKRYNSHGVREGICLTLELHAINTLCHETDHSHTPSLOFLCTISLAYTEAGOT 780
QY 781 VININGICVDTIDWMAAOPRSDGEGGQGLIKTVKLAFSYTNVIRLKPSNVSP 840
DB 781 VININGICVDTIDWMAAOPRSDGEGGQGLIKTVKLAFSYTNVIRLKPSNVSP 840
QY 841 LEQALSOHGAHNNLIATLAKYIYHKHDPALPRLAICLKRATVAPMSVYACIGNDAAA 900
DB 841 LEQALSOHGAHNNLIATLAKYIYHKHDPALPRLAICLKRATVAPMSVYACIGNDAAA 900
QY 901 IRDAFLTRLOSKIEDMRKIMLLEFLVAVETOPGLEFLNLEVKDSDGSKFESIGMW 960
DB 901 IRDAFLTRLOSKIEDMRKIMLLEFLVAVETOPGLEFLNLEVKDSDGSKFESIGMW 960
QY 961 SCILAVLIDSDQODRYWCPLLRHRAIALHLMODRDSAMVLRTRKPEMENTTSP 1020
DB 961 SCILAVLIDSDQODRYWCPLLRHRAIALHLMODRDSAMVLRTRKPEMENTTSP 1020
QY 1021 LFGTLPSPSEBSIETCALIMKIICLETIYYVKSGLDLSLDTLTKKSIERKPYMS 1080
DB 1021 LFGTLPSPSEBSIETCALIMKIICLETIYYVKSGLDLSLDTLTKKSIERKPYMS 1080
QY 1081 GYVKSALVHVAETEGSSCTSLLEYOMLVSAVRMLLIATTHADIMHLDTSVVRQLFLDY 1140
DB 1081 GYVKSALVHVAETEGSSCTSLLEYOMLVSAVRMLLIATTHADIMHLDTSVVRQLFLDY 1140
QY 1141 LDGKALLVPASVNCRLGSKMCTLLILIRKOKRELSGVDLGLPTELEBVLOADQ 1200
DB 1141 LDGKALLVPASVNCRLGSKMCTLLILIRKOKRELSGVDLGLPTELEBVLOADQ 1200
QY 1201 QLMKRTAKYFSAITYLQMKEMKVSIPQYSQVLVNCETLOEVALFQOTSHSLAG 1260
DB 1201 QLMKRTAKYFSAITYLQMKEMKVSIPQYSQVLVNCETLOEVALFQOTSHSLAG 1260
QY 1261 SATEDKDSMETDSCSRSRHRDQDGVCLGLHAKELCEVDEGDSWLYQTRRLPILPTL 1320
DB 1261 SATEDKDSMETDSCSRSRHRDQDGVCLGLHAKELCEVDEGDSWLYQTRRLPILPTL 1320
QY 1321 LTTLEVLBRKONHFEATLHLLTLARTOOGATANAAGITOSTCLPILSYVQLSTNG 1380
DB 1321 LTTLEVLBRKONHFEATLHLLTLARTOOGATANAAGITOSTCLPILSYVQLSTNG 1380
QY 1381 TAQTPSARSLDAPSWPGVYRLSMSIMEQLIKTLRYNFLPEALDFGVGHOERTLOCLNA 1440
DB 1381 TAQTPSARSLDAPSWPGVYRLSMSIMEQLIKTLRYNFLPEALDFGVGHOERTLOCLNA 1440
QY 1441 VRTYQSLACLEADHTVGTIQLSNFKEMWHFHLPOLMRDIOQVNLGYLCACTSLHSRK 1500
DB 1441 VRTYQSLACLEADHTVGTIQLSNFKEMWHFHLPOLMRDIOQVNLGYLCACTSLHSRK 1500
QY 1501 MLOHYLONKKGDDGPSAIVORVORPASAASAPSSKOPADTEASQOALHTVOGLK 1560
DB 1501 MLOHYLONKKGDDGPSAIVORVORPASAASAPSSKOPADTEASQOALHTVOGLK 1560
QY 1561 ILKXTLALAHNFPDVCOLILDSLDLAEVNFALSFITPTFDSVAPSPFGTILATVNY 1620
DB 1561 ILKXTLALAHNFPDVCOLILDSLDLAEVNFALSFITPTFDSVAPSPFGTILATVNY 1620
QY 1621 ALNMNLELDKKKEPLTOAVGLSTOABGTRTLKSLMFTMENCFYLLISQAMRYLRDPVH 1680

```

```

DB 1621 ALNMNLELDKKKEPLTOAVGLSTOABGTRTLKSLMFTMENCFYLLISQAMRYLRDPVH 1680
QY 1681 PRDKORMKOELSESLTSLSSLSRYFRRGAPSSPATGVLPSPGKSTLSKASPEOEP 1740
DB 1681 PRDKORMKOELSESLTSLSSLSRYFRRGAPSSPATGVLPSPGKSTLSKASPEOEP 1740
QY 1741 IOIQAFAVRHMOR 1753
DB 1741 IOIQAFAVRHMOR 1753

RESULT 17
US-10-719-385-19
; Sequence 19, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-19

Query Match 99.8%; Score 8991; DB 17; Length 1753;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1750; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MIRKSKITSVLSFCRSRREIMTLLGRSALREISOIEABINRWRLLEGSLTYKPPSPS 60
DB 1 MIRKSKITSVLSFCRSRREIMTLLGRSALREISOIEABINRWRLLEGSLTYKPPSPS 60
QY 61 SAEKXANKDVASPLKELGRLSKFGLDEBEGSVQLLOCYLOEDYGTBDSVTVYQDER 120
DB 61 SAEKXANKDVASPLKELGRLSKFGLDEBEGSVQLLOCYLOEDYGTBDSVTVYQDER 120
QY 121 QSOALILKADYIYBERTCILRCVHLITYFODERHPYREYVADCVKLEKELVSKYRQ 180
DB 121 QSOALILKADYIYBERTCILRCVHLITYFODERHPYREYVADCVKLEKELVSKYRQ 180
QY 181 FEELYTEAPTWETHGNLMTEROVSRWFOCLREOSMLLEIFLYAYAFEMASDILLVT 240
DB 181 FEELYTEAPTWETHGNLMTEROVSRWFOCLREOSMLLEIFLYAYAFEMASDILLVT 240
QY 241 KMFKEGFGSRQTNRLVDETMDPVDRIGYSALILVEMDIESLHKALDDRREIHP 300
DB 241 KMFKEGFGSRQTNRLVDETMDPVDRIGYSALILVEMDIESLHKALDDRREIHP 300
QY 301 AODGLICOMDCLMTFFGDI PHHAPVLLAMALLRHTLNEBETSSVYRKIGTAIQLVNQ 360
DB 301 AODGLICOMDCLMTFFGDI PHHAPVLLAMALLRHTLNEBETSSVYRKIGTAIQLVNQ 360
QY 361 YLRLQSLASGNDCTSTACMCVYGLSFVLTSELTIGNOODIITPACGVADPBL 420
DB 361 YLRLQSLASGNDCTSTACMCVYGLSFVLTSELTIGNOODIITPACGVADPBL 420
QY 421 PELFWGTBETPSGLIILDSVCGMFPHLSPILQLALVSGKSTAKVYSFLDKMSFYNE 480
DB 421 PELFWGTBETPSGLIILDSVCGMFPHLSPILQLALVSGKSTAKVYSFLDKMSFYNE 480
QY 481 LYKXKHDVISHEDGTLMRQTPKLIYPIGCGTGNLRIPOGTYGQVNLDRAYLVMEYSY 540
DB 481 LYKXKHDVISHEDGTLMRQTPKLIYPIGCGTGNLRIPOGTYGQVNLDRAYLVMEYSY 540

```

```

QY 541 SSMTLFTCEIMLHVSTADVIQHCORVXPIIDLHVXVINDLSIADCLPITSRIYML 600
Db 541 SSMTLFTCEIMLHVSTADVIQHCORVXPIIDLHVXVINDLSIADCLPITSRIYML 600
QY 601 LQRLTVISPPVDVIASCVNCLTVLAARNPAKWTDLRHGTFLPFAHVPVSLSGMISAE 660
Db 601 LQRLTVISPPVDVIASCVNCLTVLAARNPAKWTDLRHGTFLPFAHVPVSLSGMISAE 660
QY 661 GNNAGGYNLMSBOPOGEYGVITAFRLITTLVKGOLSTOSQGLVPCVMFVLEMLP 720
Db 661 GNNAGGYNLMSBOPOGEYGVITAFRLITTLVKGOLSTOSQGLVPCVMFVLEMLP 720
QY 721 SYHKRVNHSRGREGIIGLIIELTHAIINLCHETDHSHTPSIOFLICISLAYTEAGOT 780
Db 721 SYHKRVNHSRGREGIIGLIIELTHAIINLCHETDHSHTPSIOFLICISLAYTEAGOT 780
QY 781 VINIMIGVDITIDWMAAOPRSDGAGOGQOLIKTVKLAFSVTNNVIRLKPSPNVSP 840
Db 781 VINIMIGVDITIDWMAAOPRSDGAGOGQOLIKTVKLAFSVTNNVIRLKPSPNVSP 840
QY 841 LEQALSOHAGANNLIATLAKYIYKHDPALPRLAIQLKRLATVAPMSVYACIGNDAAA 900
Db 841 LEQALSOHAGANNLIATLAKYIYKHDPALPRLAIQLKRLATVAPMSVYACIGNDAAA 900
QY 901 IRDAFLRLQSKIEDMRHKWILRELTVAVEPGLIELFNLKYGQSDSKESKELGMV 960
Db 901 IRDAFLRLQSKIEDMRHKWILRELTVAVEPGLIELFNLKYGQSDSKESKELGMV 960
QY 961 SCLHAEVLIDSOQODRYWCPLLRHAAIAFLHMLQDRDPSAMLVLRKPKFENLTS 1020
Db 961 SCLHAEVLIDSOQODRYWCPLLRHAAIAFLHMLQDRDPSAMLVLRKPKFENLTS 1020
QY 1021 LFGTLPSPSESEPSILETCALIMKICLEIYVVVKGSLDOSLKTLLKKSIEKPAWYS 1080
Db 1021 LFGTLPSPSESEPSILETCALIMKICLEIYVVVKGSLDOSLKTLLKKSIEKPAWYS 1080
QY 1081 GYVSLAHVAETBESSCTSLLEYOMLVSAARMILITTHADIMHLDVSVRRQLPFDV 1140
Db 1081 GYVSLAHVAETBESSCTSLLEYOMLVSAARMILITTHADIMHLDVSVRRQLPFDV 1140
QY 1141 LDGTAKALLVPASVNCRLGSMKCTLLIILKQWRELSYDEILGPTLEILEGVLQADQ 1200
Db 1141 LDGTAKALLVPASVNCRLGSMKCTLLIILKQWRELSYDEILGPTLEILEGVLQADQ 1200
QY 1201 QLMKTKAKVSAFITVLQMKEMKVS DIPQYSQVLVANCETLQEEVIALPQTRHSALG 1260
Db 1201 QLMKTKAKVSAFITVLQMKEMKVS DIPQYSQVLVANCETLQEEVIALPQTRHSALG 1260
QY 1261 SATEDKOSMETDDGSRSHRQRODVCYLGLHAELECEVDEDEGSMQVTRRLPITL 1320
Db 1261 SATEDKOSMETDDGSRSHRQRODVCYLGLHAELECEVDEDEGSMQVTRRLPITL 1320
QY 1321 LTTLEVSJRMKONLHFTBATLHLLTLARTOOGATAVAGAGITOSICLPLSVYOLSTNG 1380
Db 1321 LTTLEVSJRMKONLHFTBATLHLLTLARTOOGATAVAGAGITOSICLPLSVYOLSTNG 1380
QY 1381 TAQTPSASRKSILDAFSWPGYRLSMSLMEQLLKTIRYNFLPEALDFVGHQERTLOCLNA 1440
Db 1381 TAQTPSASRKSILDAFSWPGYRLSMSLMEQLLKTIRYNFLPEALDFVGHQERTLOCLNA 1440
QY 1441 VRTVQSLACEBADHTVGFILQLSNFKEMHFLPOLMRDIOVNLGYLCOACTSLHSR 1500
Db 1441 VRTVQSLACEBADHTVGFILQLSNFKEMHFLPOLMRDIOVNLGYLCOACTSLHSR 1500
QY 1501 MLQHLQKNGDGLPSAFAORVQRPSPASAPSSSKOPADTEASEOALHTVOYGLK 1560
Db 1501 MLQHLQKNGDGLPSAFAORVQRPSPASAPSSSKOPADTEASEOALHTVOYGLK 1560
QY 1561 ILSTKLALRHFTPDVQIILLDOSLDAEYNFLPALSFTPTPDESEVAPSEGTLLATVNV 1620
Db 1561 ILSTKLALRHFTPDVQIILLDOSLDAEYNFLPALSFTPTPDESEVAPSEGTLLATVNV 1620

```

```

QY 1621 ALNMLGELDKKKEPLTOAVGSTOAGERTLKSILMFTMENCYLLISOAMRYLRDPVH 1680
Db 1621 ALNMLGELDKKKEPLTOAVGSTOAGERTLKSILMFTMENCYLLISOAMRYLRDPVH 1680
QY 1681 PRDQKMKOEISSELSTLSSLSRYFRRGAPSSPATGVLPSPQKSTLSLSPSPQEP 1740
Db 1681 PRDQKMKOEISSELSTLSSLSRYFRRGAPSSPATGVLPSPQKSTLSLSPSPQEP 1740
QY 1741 IQLVQAPVRHMOR 1753
Db 1741 IQLVQAPVRHMOR 1753

RESULT 18
US-10-719-385-5
; Sequence 5, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1752
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-5

Query Match 99.8%; Score 8985.5; DB 17; Length 1752;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1751; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 1 MIRSKITSVLSFCRSSRELMTIILGRSALBLSQIEALNKHMRLLLEGISYKPPSPS 60
Db 1 MIRSKITSVLSFCRSSRELMTIILGRSALBLSQIEALNKHMRLLLEGISYKPPSPS 60
QY 61 SAEKVRANKDVASPLKEIGLRSKPLGIDEOSVOLQCYOEDRGTRDSVKTVLQDER 120
Db 61 SAEKVRANKDVASPLKEIGLRSKPLGIDEOSVOLQCYOEDRGTRDSVKTVLQDER 120
QY 121 OSQALILKIDAYYBEERTCILRCVHLHLYFQDERHPYRVEYADCVDKLEKELVSKYRQ 180
Db 121 OSQALILKIDAYYBEERTCILRCVHLHLYFQDERHPYRVEYADCVDKLEKELVSKYRQ 180
QY 180 FEELYKTEAPWETHGNLMTERQVSRWFVQCLREQSMLEIIFLYAYAFEMAPSDLVLT 239
Db 180 FEELYKTEAPWETHGNLMTERQVSRWFVQCLREQSMLEIIFLYAYAFEMAPSDLVLT 239
QY 241 KMFKEQGFSGQTRHLYDETMDPVDRIQVFSALIIIEGNDISLHICALDDRELEHOF 300
Db 241 KMFKEQGFSGQTRHLYDETMDPVDRIQVFSALIIIEGNDISLHICALDDRELEHOF 300
QY 301 AODGLICODMOCMLTFGDI PHHAPVLLAMALLRHTLNPEETS SVYRKIGTALQIANFO 360
Db 301 AODGLICODMOCMLTFGDI PHHAPVLLAMALLRHTLNPEETS SVYRKIGTALQIANFO 360
QY 360 YLTRLQSLAGAGNDCTTSTACMCYGLSLFVLSLEHTLGNODDIIDTACEVLADPSL 419
Db 360 YLTRLQSLAGAGNDCTTSTACMCYGLSLFVLSLEHTLGNODDIIDTACEVLADPSL 419
QY 421 PELFWGERTSGILGIIIDSVCGMPHILSPILQILRALVSKSTAKKYYSFLDKMSFVNE 480
Db 421 PELFWGERTSGILGIIIDSVCGMPHILSPILQILRALVSKSTAKKYYSFLDKMSFVNE 480
QY 481 LYKHKPHDVISHEDGTLRRQTPKLLYPLGGQTNLRIPQGTIVGQVMDLRAYLVWEXSY 540
Db 481 LYKHKPHDVISHEDGTLRRQTPKLLYPLGGQTNLRIPQGTIVGQVMDLRAYLVWEXSY 540

```

```

Db 480 LYKHHPDVI SHEDGTLWRROT PKLLYPLGGGTNLRIPOGTVGQVMDLDRAYLVWREYSY 539
Qy 541 SSWTLFTEIEMLHVSTADVIQHCCQVKKPIIDLVHNVISTDLSINDCLPITSRIYML 600
Db 540 SSWTLFTEIEMLHVSTADVIQHCCQVKKPIIDLVHNVISTDLSINDCLPITSRIYML 599
Qy 601 LORLTVISPPVDVIASCNCCLTVLAARNPAKVTDLNHTGFLPFVAHVSLSQMISAE 660
Db 600 LORLTVISPPVDVIASCNCCLTVLAARNPAKVTDLNHTGFLPFVAHVSLSQMISAE 659
Qy 661 GNNAGYGNLMSNSPOGEYGVTTIAPLRLITTVKQGLSTQSGVLPCVMFVKEMLP 720
Db 660 GNNAGYGNLMSNSPOGEYGVTTIAPLRLITTVKQGLSTQSGVLPCVMFVKEMLP 719
Qy 721 SYHKRVNHSQVREIGCLILIELHAIINLCHENDLSHSPSLQFLCISLATEAGOT 780
Db 720 SYHKRVNHSQVREIGCLILIELHAIINLCHENDLSHSPSLQFLCISLATEAGOT 779
Qy 781 VININGIGVDTIDWMAAQPRSDGAGGQGLIKITVKLAFSVTNNVIRLKPSNVVSP 840
Db 780 VININGIGVDTIDWMAAQPRSDGAGGQGLIKITVKLAFSVTNNVIRLKPSNVVSP 839
Qy 841 LEQALSOHAGNNIIAVLAKYIYKHPDLPRLAIQULKRLATVAPMSVYACIGNDAAA 900
Db 840 LEQALSOHAGNNIIAVLAKYIYKHPDLPRLAIQULKRLATVAPMSVYACIGNDAAA 899
Qy 901 IRDALVTLQSKIEEMRIKWIIEFLVAVETOPELIFLNLKXGSDSDSKESLGMW 960
Db 900 IRDALVTLQSKIEEMRIKWIIEFLVAVETOPELIFLNLKXGSDSDSKESLGMW 959
Qy 961 SCIAHVLIELDSQODRYWCPEPLHRAAIAFLHALMODRDSAMLVNTRKPFENLTS 1020
Db 960 SCIAHVLIELDSQODRYWCPEPLHRAAIAFLHALMODRDSAMLVNTRKPFENLTS 1019
Qy 1021 LFGTLSPSEIETSEPILETALIMKIICLEIYVYVKSGLDSGLKDTLKKFSIEKPAWYS 1080
Db 1020 LFGTLSPSEIETSEPILETALIMKIICLEIYVYVKSGLDSGLKDTLKKFSIEKPAWYS 1079
Qy 1081 GYVKSIAHVATBESSCTSLLEYQMLVSAWMLIINTHADIMHLDVSVRRLPFDV 1140
Db 1080 GYVKSIAHVATBESSCTSLLEYQMLVSAWMLIINTHADIMHLDVSVRRLPFDV 1139
Qy 1141 LDGTALLLVPAVAVCLRLGSKCTLLIILRQMKRELSVDEILGPTLEILEGLQADQ 1200
Db 1140 LDGTALLLVPAVAVCLRLGSKCTLLIILRQMKRELSVDEILGPTLEILEGLQADQ 1199
Qy 1201 QLMERTKAKVPSAFTTVIOMKEMKVS DIPQYSQVLVANCETIQEEVIALFPQTRHSLAG 1260
Db 1200 QLMERTKAKVPSAFTTVIOMKEMKVS DIPQYSQVLVANCETIQEEVIALFPQTRHSLAG 1259
Qy 1261 SATETKDSWETDDCSRSRHRDQDVCVGLHAKELCEVDEGDGSMLOVTRRLPILPTL 1320
Db 1260 SATETKDSWETDDCSRSRHRDQDVCVGLHAKELCEVDEGDGSMLOVTRRLPILPTL 1319
Qy 1321 LTTLEVSILRMKONLHFTTEATLHLLTLARTOOGATAVAVAGITQSIICPLLSVYOLSTNG 1380
Db 1320 LTTLEVSILRMKONLHFTTEATLHLLTLARTOOGATAVAVAGITQSIICPLLSVYOLSTNG 1379
Qy 1381 TAQTPSASRKSIDAPSPGCVYRLSWSLMEQLIKTIRYNFLPALDPVGVHQRRTIOCLANA 1440
Db 1380 TAQTPSASRKSIDAPSPGCVYRLSWSLMEQLIKTIRYNFLPALDPVGVHQRRTIOCLANA 1439
Qy 1441 VRTVOSIACLEADHNVGFIILQSNPMKEMHFLPOLMMDIOVNLGYLCOACTSLHSHRK 1500
Db 1440 VRTVOSIACLEADHNVGFIILQSNPMKEMHFLPOLMMDIOVNLGYLCOACTSLHSHRK 1499
Qy 1501 MLQHYLQNKNGDGLSVAVAQRYORPSPASASAPSSSKQPADTEASEQALHTVOYGLX 1560
Db 1500 MLQHYLQNKNGDGLSVAVAQRYORPSPASASAPSSSKQPADTEASEQALHTVOYGLX 1559
Qy 1561 ILSKTLAALRHFTTPDVCOILLDOSLDLAEYNFLPALSTFTTPDSEVAPSGTLLATVAV 1620
Db 1560 ILSKTLAALRHFTTPDVCOILLDOSLDLAEYNFLPALSTFTTPDSEVAPSGTLLATVAV 1619

```

```

Qy 1621 ALNMLGELDKKKEPLTOAVGLSTOABGTRTLKSLMFTMENCYLLISQAMRYLRDPAV 1680
Db 1620 ALNMLGELDKKKEPLTOAVGLSTOABGTRTLKSLMFTMENCYLLISQAMRYLRDPAV 1679
Qy 1681 PRDKORMKOELSELSTLSSLSRYPFRGAPSSDPATCVLPSPQKSTSLSKASPESEGPL 1740
Db 1680 PRDKORMKOELSELSTLSSLSRYPFRGAPSSDPATCVLPSPQKSTSLSKASPESEGPL 1739
Qy 1741 IOLVOAFVRHMOR 1753
Db 1740 IOLVOAFVRHMOR 1752

RESULT 19
US-10-719-385-21
; Sequence 21, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1745
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-21

Query Match 99.2%; Score 8939; DB 17; Length 1745;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1739; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 RSSREMTILLGRALRELSQIEBELNKHWRLLLEGISYKKPPSSAEKYKANKDVASP 74
Db 7 RSSREMTILLGRALRELSQIEBELNKHWRLLLEGISYKKPPSSAEKYKANKDVASP 66
Qy 75 LKEIGLRIKSKPLGDEDSQVOLLQCYLOEDYRGTRDSVKTVLQDEROSQALILKXIADYYY 134
Db 67 LKEIGLRIKSKPLGDEDSQVOLLQCYLOEDYRGTRDSVKTVLQDEROSQALILKXIADYYY 126
Qy 135 EERTCIIACVLIHLLTYFQDEBHPYRVEYADCVDKLEKELVSKYRQOFEBELYKTEAPTWET 194
Db 127 EERTCIIACVLIHLLTYFQDEBHPYRVEYADCVDKLEKELVSKYRQOFEBELYKTEAPTWET 186
Qy 195 HGNLMTROVSRMNVQCLARESMLEIIFLYAAFEWAPSLVLTWPKFGSGSROTN 254
Db 187 HGNLMTROVSRMNVQCLARESMLEIIFLYAAFEWAPSLVLTWPKFGSGSROTN 246
Qy 255 RHLVDETMDPEVDRIGYFSALILVEGMDIESLHKCALDRERELHFOADGLICODMDCIM 314
Db 247 RHLVDETMDPEVDRIGYFSALILVEGMDIESLHKCALDRERELHFOADGLICODMDCIM 306
Qy 315 LTFGDIPIHHAVALIAMALLRHTLNPBETSSVYRKIGGTAIOANFOYLTRILQSLASGN 374
Db 307 LTFGDIPIHHAVALIAMALLRHTLNPBETSSVYRKIGGTAIOANFOYLTRILQSLASGN 366
Qy 375 DCTSTACMCYGLLSFULTSLEHTITGNODIIDTACEVADPSLPFLFMTGPTSGIG 434
Db 367 DCTSTACMCYGLLSFULTSLEHTITGNODIIDTACEVADPSLPFLFMTGPTSGIG 426
Qy 435 IILSVCGMFPHLSPILLQALRALVSGKSTAKKYSFLLDKMSFYNELYKHPHDVISHED 494
Db 427 IILSVCGMFPHLSPILLQALRALVSGKSTAKKYSFLLDKMSFYNELYKHPHDVISHED 486
Qy 495 GTLWRROT PKLLYPLGGGTNLRIPOGTVGQVMDLDRAYLVWREYSYSSWTLFTEIEML 554

```

```

487 GTWRQCPKLLPLGSGOTNLRIPGTVGQWLDRAILVREYSSWTLFTGIEML 546
QY HVSTADAVIQHCQRVKPIIDLVHKVISTDLSTADCLPITSRIYMLQRLTTVISPV 614
Db HVSSTADVIQHCQRVKPIIDLVHKVISTDLSTADCLPITSRIYMLQRLTTVISPV 606
QY IASCVNCLTVLAARBPATVMDLRTGTGLPFAHVPSSLSQWISAEKNAGGYGLMNS 674
Db IASCVNCLTVLAARBPATVMDLRTGTGLPFAHVPSSLSQWISAEKNAGGYGLMNS 666
QY EPOGEYVTAIFLRLITTLVKGQSGTOSQGLVPCWFLKEMPSYHKRYSNGVRE 734
Db EPOGEYVTAIFLRLITTLVKGQSGTOSQGLVPCWFLKEMPSYHKRYSNGVRE 726
QY QIGCLILBLIHAAILNLCHETDLHSSHTPSLOFLCISLAYTEAGTVINIGIVDTID 794
Db QIGCLILBLIHAAILNLCHETDLHSSHTPSLOFLCISLAYTEAGTVINIGIVDTID 786
QY VMAAPRSDGAGGQGGOLIKTVLAFSVNNVRLRPPSVNVSPLBOALSQGHAGNN 854
Db VMAAPRSDGAGGQGGOLIKTVLAFSVNNVRLRPPSVNVSPLBOALSQGHAGNN 846
QY LIAVLAKYIYKHDPALRLAIQLKRLATVAPMSVYACLGNDAAIRDAFLTRLOSKE 914
Db LIAVLAKYIYKHDPALRLAIQLKRLATVAPMSVYACLGNDAAIRDAFLTRLOSKE 906
QY DMRIKVMLEPLTVAVETOPGLIELFLNLVKGSDGSKESPGLGWSCLHAVALLEIDSQ 974
Db DMRIKVMLEPLTVAVETOPGLIELFLNLVKGSDGSKESPGLGWSCLHAVALLEIDSQ 966
QY ODRVWCPPLHRAALAFIHALMODRDSAMVLRTKPKFWELTSPRLGCTLSPESTSE 1034
Db ODRVWCPPLHRAALAFIHALMODRDSAMVLRTKPKFWELTSPRLGCTLSPESTSE 1026
QY QDRVWCPPLHRAALAFIHALMODRDSAMVLRTKPKFWELTSPRLGCTLSPESTSE 1026
Db QDRVWCPPLHRAALAFIHALMODRDSAMVLRTKPKFWELTSPRLGCTLSPESTSE 1026
QY SILETCALIMKICLEIYVYVKSGLDOSLKTOKKFSIEKRAYSGVYKSLAVNAETE 1094
Db SILETCALIMKICLEIYVYVKSGLDOSLKTOKKFSIEKRAYSGVYKSLAVNAETE 1086
QY SILETCALIMKICLEIYVYVKSGLDOSLKTOKKFSIEKRAYSGVYKSLAVNAETE 1086
Db SILETCALIMKICLEIYVYVKSGLDOSLKTOKKFSIEKRAYSGVYKSLAVNAETE 1086
QY GSSCTSLLEYOMLVASAMRLIIATTHADIMHLDTSVVRQGLFLVDLGTALLLVPASV 1154
Db GSSCTSLLEYOMLVASAMRLIIATTHADIMHLDTSVVRQGLFLVDLGTALLLVPASV 1146
QY NCLRIGSKKCTLLILRQMKRELGSVDEILGPTLEILBGLVQADQOMKTKAVVSAF 1214
Db NCLRIGSKKCTLLILRQMKRELGSVDEILGPTLEILBGLVQADQOMKTKAVVSAF 1206
QY NCLRIGSKKCTLLILRQMKRELGSVDEILGPTLEILBGLVQADQOMKTKAVVSAF 1206
Db NCLRIGSKKCTLLILRQMKRELGSVDEILGPTLEILBGLVQADQOMKTKAVVSAF 1206
QY ITVLQMKEMKYSIDIPQYSOLVNVCEETOEBEYIALFDOTRSHLAGSATBEDKSMETDC 1274
Db ITVLQMKEMKYSIDIPQYSOLVNVCEETOEBEYIALFDOTRSHLAGSATBEDKSMETDC 1266
QY ITVLQMKEMKYSIDIPQYSOLVNVCEETOEBEYIALFDOTRSHLAGSATBEDKSMETDC 1266
Db ITVLQMKEMKYSIDIPQYSOLVNVCEETOEBEYIALFDOTRSHLAGSATBEDKSMETDC 1266
QY SRSRRRDQDVCVGLHAKELCEVDEBDGSMLOVTRRLPILPTLLTTLVSLMKONL 1334
Db SRSRRRDQDVCVGLHAKELCEVDEBDGSMLOVTRRLPILPTLLTTLVSLMKONL 1326
QY SRSRRRDQDVCVGLHAKELCEVDEBDGSMLOVTRRLPILPTLLTTLVSLMKONL 1334
Db SRSRRRDQDVCVGLHAKELCEVDEBDGSMLOVTRRLPILPTLLTTLVSLMKONL 1326
QY HFTETTLHLTLTATTOGATVAVAGITOSTCLPILSVYQUSSTMGTAOTPSAKSIDA 1394
Db HFTETTLHLTLTATTOGATVAVAGITOSTCLPILSVYQUSSTMGTAOTPSAKSIDA 1386
QY HFTETTLHLTLTATTOGATVAVAGITOSTCLPILSVYQUSSTMGTAOTPSAKSIDA 1386
Db HFTETTLHLTLTATTOGATVAVAGITOSTCLPILSVYQUSSTMGTAOTPSAKSIDA 1386
QY PSMPEVYVLSMSLMEQLKTLRYNFLPEALDFVGHOBERTLOCLNAVTVOSIACLEBAD 1454
Db PSMPEVYVLSMSLMEQLKTLRYNFLPEALDFVGHOBERTLOCLNAVTVOSIACLEBAD 1446
QY PSMPEVYVLSMSLMEQLKTLRYNFLPEALDFVGHOBERTLOCLNAVTVOSIACLEBAD 1454
Db PSMPEVYVLSMSLMEQLKTLRYNFLPEALDFVGHOBERTLOCLNAVTVOSIACLEBAD 1446
QY HTVGFILOLSNPMKEMHFLPLQMDIQVNLGYLCOACTSLHSHKMLQHYLQNRNGDL 1514
Db HTVGFILOLSNPMKEMHFLPLQMDIQVNLGYLCOACTSLHSHKMLQHYLQNRNGDL 1506
QY HTVGFILOLSNPMKEMHFLPLQMDIQVNLGYLCOACTSLHSHKMLQHYLQNRNGDL 1514
Db HTVGFILOLSNPMKEMHFLPLQMDIQVNLGYLCOACTSLHSHKMLQHYLQNRNGDL 1506
QY PSANVQRVQRPSSASAPSSSKOPADTEASBOQALHTVOYGLKILSKTLAARHFTP 1574
Db PSANVQRVQRPSSASAPSSSKOPADTEASBOQALHTVOYGLKILSKTLAARHFTP 1566
QY PSANVQRVQRPSSASAPSSSKOPADTEASBOQALHTVOYGLKILSKTLAARHFTP 1574
Db PSANVQRVQRPSSASAPSSSKOPADTEASBOQALHTVOYGLKILSKTLAARHFTP 1566
QY DVCCQLLDQSLDLAEYNFLFALSFTTPTFDSVAPSGTILATVVALNMLGELDKKEP 1634
Db DVCCQLLDQSLDLAEYNFLFALSFTTPTFDSVAPSGTILATVVALNMLGELDKKEP 1634

```

```

Db DVCCQLLDQSLDLAEYNFLFALSFTTPTFDSVAPSGTILATVVALNMLGELDKKEP 1626
QY LTOAVGLSTQAEGRITLKSLLMFTMENCYLLISQANRYLADPAVHPDKQRMQELSS 1694
Db LTOAVGLSTQAEGRITLKSLLMFTMENCYLLISQANRYLADPAVHPDKQRMQELSS 1686
QY LSTLLSLSRFRGAPSPATGVLPSPQGSTLSKSPESOEPLQLOVAFVHMOR 1753
Db LSTLLSLSRFRGAPSPATGVLPSPQGSTLSKSPESOEPLQLOVAFVHMOR 1745

RESULT 20
US-10-370-715B-544
; Sequence 544, Application US/10370715B
; Publication No. US20040258678A1
; GENERAL INFORMATION:
; Patin Docket Preview
; APPLICANT: BODARY, SARAH C.
; APPLICANT: CLARK, HILARY
; APPLICANT: BRISDELL, HUNTER
; APPLICANT: JACKMAN, JONET
; APPLICANT: SCHOENFELD, JILL R.
; APPLICANT: WILLIAMS, P. MICKEY
; APPLICANT: WOOD, WILLIAM I.
; APPLICANT: MU, THOMAS D.
; TITLE OF INVENTION: Compositions and Methods for the Treatment of Immune
; FILE REFERENCE: P1948R1-US
; CURRENT FILING DATE: 2003-02-21
; NUMBER OF SEQ ID NOS: 742
; SEQ ID NO 544
; LENGTH: 1745
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-370-715B-544

Query Match 99.2%; Score 8939; DB 17; Length 1745;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1739; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 RSSRELWTLLGRSALRELSQIEBAELNGHMRLLLEGSLYKPPSPSSAEKYANKDVAS 74
Db 7 RSSRELWTLLGRSALRELSQIEBAELNGHMRLLLEGSLYKPPSPSSAEKYANKDVAS 66
QY 75 LKEGLRISKFLGLDEBQSVLLQCYLOEDYRGRDSVKTVALQDERQSOALLIKIADYY 134
Db 67 LKEGLRISKFLGLDEBQSVLLQCYLOEDYRGRDSVKTVALQDERQSOALLIKIADYY 126
QY 135 EERTCILACVHLTYFDDEHNPRAVEADCVDLKELVSKYKQOFELYKTAFTMET 194
Db 127 EERTCILACVHLTYFDDEHNPRAVEADCVDLKELVSKYKQOFELYKTAFTMET 186
QY 195 HGNLMTQVSRMFWQCLREOSMLLEIFLYAAYFEMAPSLVLTAKFKEQSGFSRQTN 254
Db 187 HGNLMTQVSRMFWQCLREOSMLLEIFLYAAYFEMAPSLVLTAKFKEQSGFSRQTN 246
QY 255 RHLVDETMDPVDRIGYFSALILVEGMDIESLHKCALDRRELHQFADGILCQMDCLM 314
Db 247 RHLVDETMDPVDRIGYFSALILVEGMDIESLHKCALDRRELHQFADGILCQMDCLM 306
QY 315 LTFQGIPIHRAVLLAMALLBHTLNPBETSSVVRKICGTAIQLVNFQVYTRLLQSLASGN 374
Db 307 LTFQGIPIHRAVLLAMALLBHTLNPBETSSVVRKICGTAIQLVNFQVYTRLLQSLASGN 366
QY 375 DCTSTACMCYGLSLFVLTSLLEHTTGNQODIDTACEVLADPSLPFLFWGTBPTSGLG 434
Db 367 DCTSTACMCYGLSLFVLTSLLEHTTGNQODIDTACEVLADPSLPFLFWGTBPTSGLG 426
QY 435 IILDSVCGMFPHLSPILLQILRALVSGSKTAKKYSFLDKXGFPNELYKHPHDVISHED 494
Db 427 IILDSVCGMFPHLSPILLQILRALVSGSKTAKKYSFLDKXGFPNELYKHPHDVISHED 486

```

495 GTLMRRQTPKLLYPLGGQTNLRIPOGTVGQVWLDRAVLVRWEYSWSMTLFTCEIEMLL 554  
487 GTLMRRQTPKLLYPLGGQTNLRIPOGTVGQVWLDRAVLVRWEYSWSMTLFTCEIEMLL 546  
555 HVAVSADVIQHCORVXPIIDLVHXYISTDLSTADCLLPITSRIYMLQRLTTVISPPVDV 614  
547 HVAVSADVIQHCORVXPIIDLVHXYISTDLSTADCLLPITSRIYMLQRLTTVISPPVDV 606  
615 IASCNCLTVLAARPAKVTMDLRHTGFLPFVAHPVSSLSOMISAEAGNAGGYGLMNS 674  
607 IASCNCLTVLAARPAKVTMDLRHTGFLPFVAHPVSSLSOMISAEAGNAGGYGLMNS 666  
675 EOPQGEYGVTAFLRLITLVKQGLSTQSGVLPCVAFVLEKMLPSYHKRKYNSHGYRE 734  
667 EOPQGEYGVTAFLRLITLVKQGLSTQSGVLPCVAFVLEKMLPSYHKRKYNSHGYRE 726  
735 QIGCIIILBIHAIIIMLCHETDLSHSHTSLOPLCISLAIYBAGQTVINIGIGVDTIDM 794  
727 QIGCIIILBIHAIIIMLCHETDLSHSHTSLOPLCISLAIYBAGQTVINIGIGVDTIDM 786  
795 VMAAPRSDGAEQGGQGLIKTVLAFSVTNVRIKLPSPNVSPLEQALSOHAGHNN 854  
787 VMAAPRSDGAEQGGQGLIKTVLAFSVTNVRIKLPSPNVSPLEQALSOHAGHNN 846  
855 LIAVLAKYIYHKGDPALPRLAIQILKRLATVAPMSVYACLGNDAAIRDAEFLTRLOSXIE 914  
847 LIAVLAKYIYHKGDPALPRLAIQILKRLATVAPMSVYACLGNDAAIRDAEFLTRLOSXIE 906  
915 DMRIKVMLEFLTVAVETOPGLIELFNLLEVYDGSKEPSLGWSCGLAVVBLIDSOQ 974  
907 DMRIKVMLEFLTVAVETOPGLIELFNLLEVYDGSKEPSLGWSCGLAVVBLIDSOQ 966  
975 QDRYWCPLLRRAAIAFLHALMODRDSAMLVLRTPKPFMENLTPSLFGTSPSETSEP 1034  
967 QDRYWCPLLRRAAIAFLHALMODRDSAMLVLRTPKPFMENLTPSLFGTSPSETSEP 1026  
1035 SILETCALIMKICIEIYVYVKSIDQSLKDTLKKFESTERKPAYWSGYKSLAVHVAETE 1094  
1027 SILETCALIMKICIEIYVYVKSIDQSLKDTLKKFESTERKPAYWSGYKSLAVHVAETE 1086  
1095 GSSCSTSLLEYQMLVAMRMLIINTHADIMHLDVSVRROLFLVDGTALLVPAVS 1154  
1087 GSSCSTSLLEYQMLVAMRMLIINTHADIMHLDVSVRROLFLVDGTALLVPAVS 1146  
1155 NCLRISGSKCTLLIILRQMKRELSVDEILGPLTEILEGYLOADQOLMEKTKAXVSFA 1214  
1147 NCLRISGSKCTLLIILRQMKRELSVDEILGPLTEILEGYLOADQOLMEKTKAXVSFA 1206  
1215 ITVYQMKEMKYSDIPOYSQVLVANCETIOEVIALFDQTRHSLAGSATEDKXMETDDC 1274  
1207 ITVYQMKEMKYSDIPOYSQVLVANCETIOEVIALFDQTRHSLAGSATEDKXMETDDC 1266  
1275 SRSRHRDQDGVYGLHLAKELCEVDEDEGSMLOVTRRLPILPLTLTTEVSLRMKNL 1334  
1267 SRSRHRDQDGVYGLHLAKELCEVDEDEGSMLOVTRRLPILPLTLTTEVSLRMKNL 1326  
1335 HFTFATLHLLTLTARQOGATAVAGITQISICLPLSVYOLSTNGTAQTPSASRKSIDA 1394  
1327 HFTFATLHLLTLTARQOGATAVAGITQISICLPLSVYOLSTNGTAQTPSASRKSIDA 1386  
1395 PSWNGVYRLSMLBOLIKTIRYVFLPALDPVGHQERTLOCNNAVTVQSLACLEAD 1454  
1387 PSWNGVYRLSMLBOLIKTIRYVFLPALDPVGHQERTLOCNNAVTVQSLACLEAD 1446  
1455 HTVGFIIQLSNFKEMHFLPOLMRDIOVNLGYCOACTSLHSRKMLOHLOKNDG 1514  
1447 HTVGFIIQLSNFKEMHFLPOLMRDIOVNLGYCOACTSLHSRKMLOHLOKNDG 1506  
1515 PSAAVQVQRPSPASASAPSSSKOPADTEASEOQALHTVQYGLKILSKTLAALRHFTP 1574  
1507 PSAAVQVQRPSPASASAPSSSKOPADTEASEOQALHTVQYGLKILSKTLAALRHFTP 1566  
1575 DVQOILDDQSLDLAEVNFPLFALSTFTTTPDSEVAPSGTILATVNALMGLDKKKEP 1634

1567 DVQOILDDQSLDLAEVNFPLFALSTFTTTPDSEVAPSGTILATVNALMGLDKKKEP 1626  
1635 LTOAVGLSTQAEGRITLKSILMTWENCYFILLISQAMRYRDPVAPRDKQRMKOLSE 1694  
1627 LTOAVGLSTQAEGRITLKSILMTWENCYFILLISQAMRYRDPVAPRDKQRMKOLSE 1686  
1695 LSTLSSLSRYFRFGAPSPATGVLPSPOKSTLSASPSOEPILQLOVAPFRMHOR 1753  
1687 LSTLSSLSRYFRFGAPSPATGVLPSPOKSTLSASPSOEPILQLOVAPFRMHOR 1745

RESULT 21  
US-10-719-385-22  
; Sequence 22, Application US/10719385  
; Publication No. US20040209284A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Toole et al.  
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
; FILE REFERENCE: 22058-582  
; CURRENT APPLICATION NUMBER: US/10/719,385  
; PRIOR FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: PCT/US03/37339  
; PRIOR FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: 60/428,094  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 22  
; LENGTH: 3534  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-719-385-22

Query Match 99.1%; Score 8927; DB 17; Length 3534;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1739; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

15 RSSRELITLGRSALREISQIEAEINKHWRRLLEGLSYKPPSPSAEKYKANKOVASP 74  
25 RSSRELITLGRSALREISQIEAEINKHWRRLLEGLSYKPPSPSAEKYKANKOVASP 84  
75 LKEIGLRISKFLGLDEEVSQVLLQCYLOEDYRGTDSVKTVALDEROSQALILKIDYVY 134  
85 LKEIGLRISKFLGLDEEVSQVLLQCYLOEDYRGTDSVKTVALDEROSQALILKIDYVY 144  
135 EERTCTICRCLHLLTYFODERHRYRYEVADCVNKELEKVSRTQOPEELYKTEAPTWET 194  
145 EERTCTICRCLHLLTYFODERHRYRYEVADCVNKELEKVSRTQOPEELYKTEAPTWET 204  
195 HGNLMTERRQVSRMFVOCLEBSQWMLLEIIFLYAYVFEKAPSDLLVLTMPKEGFGSRQTN 254  
205 HGNLMTERRQVSRMFVOCLEBSQWMLLEIIFLYAYVFEKAPSDLLVLTMPKEGFGSRQTN 264  
255 RHLVDETMDPFVDRIGYFSALIIIVEGMDIESLHKCALDDBRELIHQAPDGLICQMDCLM 314  
265 RHLVDETMDPFVDRIGYFSALIIIVEGMDIESLHKCALDDBRELIHQAPDGLICQMDCLM 324  
315 LTRGDIPIHAPVLLAVALRHTLNPEETSIVYKIGTGAIQLVNFOYLTRLLQSLASGN 384  
325 LTRGDIPIHAPVLLAVALRHTLNPEETSIVYKIGTGAIQLVNFOYLTRLLQSLASGN 384  
375 DCTTSTACMCVYGLSFLVITSLEHLTGNQODIIPRACEVADPSLPFLFWGTEPTSGIG 434  
385 DCTTSTACMCVYGLSFLVITSLEHLTGNQODIIPRACEVADPSLPFLFWGTEPTSGIG 444  
435 IILDVCGMFPHLISPLQLLRALVSGKSTAKKVSFLDKMSFYNELYGRKPDVISHED 494  
445 IILDVCGMFPHLISPLQLLRALVSGKSTAKKVSFLDKMSFYNELYGRKPDVISHED 504  
495 GTLMRRQTPKLLYPLGGQTNLRIPOGTVGQVWLDRAVLVRWEYSWSMTLFTCEIEMLL 554  
505 GTLMRRQTPKLLYPLGGQTNLRIPOGTVGQVWLDRAVLVRWEYSWSMTLFTCEIEMLL 564



```

Qy 555 HVSTADVIQHCORVKPIIDLVHKVISTDLSTADCLPITSRIYMLQRLTTVISPPDV 614
Db 565 HVSTADVIQHCORVKPIIDLVHKVISTDLSTADCLPITSRIYMLQRLTTVISPPDV 624
Qy 615 IASCVNCLTVLAARNPAKWTDLRHGTGFLPFVAHPVSSLQMSISAGNAGCGNLMNS 674
Db 625 IASCVNCLTVLAARNPAKWTDLRHGTGFLPFVAHPVSSLQMSISAGNAGCGNLMNS 684
Qy 675 EQPGEYGVTAFLRLITTLVKGQSGTOSQGLVPCWVFKEMLPSTHKRRYNSHGVR 734
Db 685 EQPGEYGVTAFLRLITTLVKGQSGTOSQGLVPCWVFKEMLPSTHKRRYNSHGVR 744
Qy 735 QIGCLIEHLIAIINLCHETDLHSHTPSLOPLICSLATYBAGQTVNINIGVDTDM 794
Db 745 QIGCLIEHLIAIINLCHETDLHSHTPSLOPLICSLATYBAGQTVNINIGVDTDM 804
Qy 795 VMAAPRSDGAEQGGQGLIKTVLAFTSVNNVIRLKPNSNVSPLEQALSHGAGNN 854
Db 805 VMAAPRSDGAEQGGQGLIKTVLAFTSVNNVIRLKPNSNVSPLEQALSHGAGNN 864
Qy 855 LIATLAKTYHKHDPALPRLAIQILKRLATVAPMSVACLGNDAAIRDAFLTRLOSKE 914
Db 865 LIATLAKTYHKHDPALPRLAIQILKRLATVAPMSVACLGNDAAIRDAFLTRLOSKE 924
Qy 915 DMRIKWMIEFLTVAVETOPGLIELFLNLEVKGDSGSKESLGWMSGLHVLVLEIDSOQ 974
Db 925 DMRIKWMIEFLTVAVETOPGLIELFLNLEVKGDSGSKESLGWMSGLHVLVLEIDSOQ 984
Qy 975 QDRYWCPEPLHRAAIAFLHALMODRDSAMLVLRTPKFWENLISPLFGTSPSETSEP 1034
Db 985 QDRYWCPEPLHRAAIAFLHALMODRDSAMLVLRTPKFWENLISPLFGTSPSETSEP 1044
Qy 1035 SILETCALIMKICIEIYVVVKGSLDOSLKDCLKKFSIEKRPAYMSGYKSLAVHVAETE 1094
Db 1045 SILETCALIMKICIEIYVVVKGSLDOSLKDCLKKFSIEKRPAYMSGYKSLAVHVAETE 1104
Qy 1095 GSSCSTSLLEYOMIVSAMRMLIIATTHADIMHLPDSVVRQLFDVLDGTALLLVPSV 1154
Db 1105 GSSCSTSLLEYOMIVSAMRMLIIATTHADIMHLPDSVVRQLFDVLDGTALLLVPSV 1164
Qy 1155 NCLRISGSKCTLLILLRQMK----RELGSYDEILGPLETEILEGVLQADQOLMEKTKAV 1210
Db 1165 NCLRISGSKCTLLILLRQMKSIISRELGSYDEILGPLETEILEGVLQADQOLMEKTKAV 1224
Qy 1211 FSAFITYLOMKEMKVSIDIPQYSQVLVANCETLOEVEVIALPDQTRHSIALGSATEDKOSME 1270
Db 1225 FSAFITYLOMKEMKVSIDIPQYSQVLVANCETLOEVEVIALPDQTRHSIALGSATEDKOSME 1284
Qy 1271 TDDGSRSHRORQDVCVLGHLAKELECEVEDGDSMWLOVTRRLPILPTLLTLEVSIRM 1330
Db 1285 TDDGSRSHRORQDVCVLGHLAKELECEVEDGDSMWLOVTRRLPILPTLLTLEVSIRM 1344
Qy 1331 KONLHFEATLHLTLTARTOOGATAVAGAGITQSCILPLSVYOLSTNGTAQTPSARX 1390
Db 1345 KONLHFEATLHLTLTARTOOGATAVAGAGITQSCILPLSVYOLSTNGTAQTPSARX 1404
Qy 1391 SLDAPSWPGVYRLSWSLMEQLKTLRYNPLPEALDPVGHQERTLOCLANAVTYQSLACL 1450
Db 1405 SLDAPSWPGVYRLSWSLMEQLKTLRYNPLPEALDPVGHQERTLOCLANAVTYQSLACL 1464
Qy 1451 BEADHTVGFIIQLSFMKEHMFHLBOLMRDIOVNLGYICOACTSILSHRRKMLQHLQONK 1510
Db 1465 BEADHTVGFIIQLSFMKEHMFHLBOLMRDIOVNLGYICOACTSILSHRRKMLQHLQONK 1524
Qy 1511 GDGPSPAAVQRPSPASASAPSSSKOPADTEASEGOALHTYOGYGLKTLISKTALNR 1570
Db 1525 GDGPSPAAVQRPSPASASAPSSSKOPADTEASEGOALHTYOGYGLKTLISKTALNR 1584
Qy 1571 HFTPDVQCIILLDOSIDLAEYNFLFALSTPTPTFDESEVAPSFQTLLATYVVALNMLGELDK 1630
Db 1585 HFTPDVQCIILLDOSIDLAEYNFLFALSTPTPTFDESEVAPSFQTLLATYVVALNMLGELDK 1644

```

```

Qy 1631 KKEPLTOAVGLSTOAEGRTRTLKSLMFTMENCFFYLLISQAMRYLRDPAVHRDKORMKOE 1690
Db 1645 KKEPLTOAVGLSTOAEGRTRTLKSLMFTMENCFFYLLISQAMRYLRDPAVHRDKORMKOE 1704
Qy 1691 LSSELSTLSSLSYFRRGAPSSPATGYLPPQCKSTSLSPASPSQEPILQLOVAFVRH 1750
Db 1705 LSSELSTLSSLSYFRRGAPSSPATGYLPPQCKSTSLSPASPSQEPILQLOVAFVRH 1764
Qy 1751 MOR 1753
Db 1765 MOR 1767

RESULT 22
US-10-719-385-23
; Sequence 23, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 1111
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-23

Query Match 58.5%; Score 5270; DB 17; Length 1111;
Best Local Similarity 92.8%; Pred. No. 0;
Matches 1031; Conservative 36; Mismatches 34; Indels 10; Gaps 1;

Qy 653 LSQMSISAGNAGGYGULNMSSEOPQGEYGVTAFLRLITTLVKGQSGTOSQGLVPCW 712
Db 1 MTQMSISAGNAGGYGULNMSSEOPQGEYGVTAFLRLITTLVKGQSGTOSQGLVPCW 60
Qy 713 FVLKEMLPSTHKRRYNSHGVELIGCLIEHLIAIINLCHETDLHSHTPSLOPLICSL 772
Db 61 FVLKEMLPSTHKRRYNSHGVELIGCLIEHLIAIINLCHETDLHSHTPSLOPLICSL 120
Qy 773 AYTEAGQTVINIMIGVDTIDMMAAOPRSDGAEQGGQGLIKTVKLAFTSVNNVIRLK 832
Db 121 AYTEAGQTVINIMIGVDTIDMMAAOPRSDGAEQGGQGLIKTVKLAFTSVNNVIRLK 180
Qy 833 PPSNVSPLEQALSQHGAGNNLIATLAKTYHKHDPALPRLAIQILKRLATVAPMSYTA 892
Db 181 PPSNVSPLEQALSQHGAGNNLIATLAKTYHKHDPALPRLAIQILKRLATVAPMSYTA 240
Qy 893 CLGNDAAIRDAFLTRLOSKEDEMRKIMLEPLTVAVETOPGLIELFLNLEVKGDSGS 952
Db 241 CLGNDAAIRDAFLTRLOSKEDEMRKIMLEPLTVAVETOPGLIELFLNLEVKGDSGS 300
Qy 953 KEPLISGMSCLHAYLELIDSOQDRCYPCPLHRAAIAFLHALMODRDSAMLVLRTPKX 1012
Db 301 KEPLISGMSCLHAYLELIDSOQDRCYPCPLHRAAIAFLHALMODRDSAMLVLRTPKX 360
Qy 1013 FWNENLSPFTGTLSPSETSPSILETCALIMKICIEIYVVVKGSLDOSLKDCLKKFSI 1072
Db 361 FWNENLSPFTGTLSPSETSPSILETCALIMKICIEIYVVVKGSLDOSLKDCLKKFSI 420
Qy 1073 EKRPAYMSGYKSLAVHVAETEGSSCTSLLEYOMIVSAMRMLIIATTHADIMHLPDSV 1132
Db 421 EKRPAYMSGYKSLAVHVAETEGSSCTSLLEYOMIVSAMRMLIIATTHADIMHLPDSV 480
Qy 1133 RROPLDVLDTGKALLVPASVNCILGSMKCTLLILLRQMKRELGSVDEILGPLETEIL 1192

```



```

Db      481 RROFLDVLDTKALLVAASVNCIRLGSNMCTLLILRQMKELGAVKILPTEIL 540
Qy      1193 EGVQAOQOQMEKXKAVFSAFITYLQMKEMKVSIDIPQYSQLVNVCTLOEEVIALFDQ 1252
Db      541 EGVQAOQOQMEKXKAVFSAFITYLQMKELRQVIGIPYSQLVNVCTLOEEVIALFDQ 600
Qy      1253 TRHSLALGSATEDSDMETDDCSRRHRDQDGVCLGLHLAKELCEVEDSDMLQVTR 1312
Db      601 TRHSLASASAEDESDMETDDCSRRHRDQDGVCLGLHLAKELCEVEDSDMLQVTR 660
Qy      1313 RLPLPILTLTLEVSLEKMKQNLHFTFATLTLTLARTQCATVAGAGITQICLPLLS 1372
Db      661 RLPLPILTLTLEVSLEKMKQNLHFTFATLTLTLARTQCATVAGAGITQICLPLLS 720
Qy      1373 VYQISTNCTACTPSASRSKSLDAPSPGYYRLSMSLMEQLKTLRYNPLPEALDVGYNQOE 1432
Db      721 VYQISSNCTGQTPSTSRKSLDAPSPGYYRLSMSLMEQLKTLRYNPLPEALDVGYNQOE 780
Qy      1433 RTLOCLINAVRTVQSLACLSEADHTVGFILQLSNFKEMHFFHLPOLMRDIOVNLGYLCOAC 1492
Db      781 RTLOCLINAVRTVQSLACLSEADHTVGFILQLSHPRKEMHFFHLPOLMRDIOVNLGYLCOAC 840
Qy      1493 TSLHSRMTLOHYLQNNKXGDLPSAVQORVORPPSA-----ASAPSSSKOPAD 1542
Db      841 TSLHSRMTLOHYLQNNKXGDLPSAVQORVORPPSA-----ASAPSSSKOPAD 900
Qy      1543 TEASEQOALHTVOYGLKILSKTLAALHFTPDVQOILLDOSLDLAEYNPLFALSFTPT 1602
Db      901 TEASEQOALHTVOYGLKILSKTLAALHFTPDVQOILLDOSLDLAEYNPLFALSFTPT 960
Qy      1603 FDSEVAPFSGTLLATVNVVALNMLGELDKKEPLTQAVGLSTQABGTRTLKSLNFTWENC 1662
Db      961 FDSEVAPFSGTLLATVNVVALNMLGELDKKESLTQAVGLSTQABGTRTLKSLNFTWENC 1020
Qy      1663 FYLLISQAMRYLRDPAVHPRDKQKQKQELSELSTLSLSRYRRGAPSSPATGVLPSP 1722
Db      1021 FYLLISQAMRYLRDPAVHPRDKQKQKQELSELSTLSLSRYRRGAPSSPATGVLPSP 1080
Qy      1723 QGKSTSLSKASPESEOPILQVQAFVRHMOR 1753
Db      1081 QGKSTSLSKASPESEOPILQVQAFVRHMOR 1111

```

## RESULT 23

```

US-10-719-385-24
; Sequence 24, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 853
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-24

```

Query Match 45.4%; Score 4093; DB 17; Length 853;

Best Local Similarity 96.2%; Pred. No. 0;

Matches 810; Conservative 5; Mismatches 21; Indels 6; Gaps 1;

Qy 683 VTIAFLRLITTLVKGQLSTQSGIIVPCVMFYLKMLPSYHKWRYSNGVREQIGCLILE 742

Db 1 VTIAFLRLITTLVKGQLSTQSGIIVPCVMFYLKMLPSYHKWRYSNGVREQIGCLILE 60

```

Qy      743 LTHAILNLCHETDLHSHSTPSLOFLCISLAYEAGQVINIMGIVDTIDMVAAQPRS 802
Db      61 LTHAILNLCHETDLHSHSTPSLOFLCISLAYEAGQVINIMGIVDTIDMVAAQPRS 120
Qy      803 DGAEGGQGGQLIKTVKLASVTNNVYRLKRPSPNVSPLEQALSGHAGNNLIAYLAKY 862
Db      121 DGAEGGQGGQLIKTVKLASVTNNVYRLKRPSPNVSPLEQALSGHAGNNLIAYLAKY 180
Qy      863 IYHKHDPALRLAIQLKRLATVAPMSVYACLGNDAAIIDAFLTRLOSKIEDRKIKMI 922
Db      181 IYHKHDPALRLAIQLKRLATVAPMSVYACLGNDAAIIDAFLTRLOSKIEDRKIKMI 240
Qy      923 LEPFLTAVETOPGLIEFLNLEVDGSDGSKESLGMWSCIHAVALBELIDSGQDQRYVCP 982
Db      241 LEPFLTAVETOPGLIEFLNLEVDGSDGSKESLGMWSCIHAVALBELIDSGQDQRYVCP 300
Qy      983 LTHRAIAFLHALMODRRDSAMVLRKPKFWMENLSPLEGTLSPEBETSEPSILETCAL 1042
Db      301 LTHRAIAFLHALMODRRDSAMVLRKPKFWMENLSPLEGTLSPEBETSEPSILETCAL 360
Qy      1043 IMKICLIEIYVVGSLDQSLKDTLKKEFSIEKRPAYSGYVKSIAVVAETEGSSCTSL 1102
Db      361 IMKICLIEIYVVGSLDQSLKDTLKKEFSIEKRPAYSGYVKSIAVVAETEGSSCTSL 420
Qy      1103 EYQMLVSAMEMLLIATTHADIMHLTDSVVRQFLDVLDTKALLVPASVNCIRLGS 1162
Db      421 EYQMLVSAMEMLLIATTHADIMHLTDSVVRQFLDVLDTKALLVPASVNCIRLGS 480
Qy      1163 KCTLLILLLQMKRELGSVDEILGPLEIIEGVQAOQOQMEKXKAVFSAFITYLQMK 1222
Db      481 KCTLLILLLQMKRELGSVDEILGPLEIIEGVQAOQOQMEKXKAVFSAFITYLQMK 540
Qy      1223 MKVSDIPQYSQLVNVCTLOEEVIALFDQTRHSLALGSATEDSDMETDDCSRRHQ 1282
Db      541 MKVSDIPQYSQLVNVCTLOEEVIALFDQTRHSLALGSATEDSDMETDDCSRRHQ 600
Qy      1283 RDGVCVLGLHLAKELCEVEDSDGSMQVTRRLPLPILTLTLEVSLEKMKQNLHFTFATLH 1342
Db      601 RDGVCVLGLHLAKELCEVEDSDGSMQVTRRLPLPILTLTLEVSLEKMKQNLHFTFATLH 660
Qy      1343 LLLTLARTQCATVAGAGITQICLPLLSYQSLSTNGTQTPSASRSKSLDAPSPGYYR 1402
Db      661 LLLTLARTQCATVAGAGITQICLPLLSYQSLSTNGTQTPSASRSKSLDAPSPGYYR 720
Qy      1403 LSNFMEQLKTLRYNPLPEALDPVGVHOBRTLOCLINAVTVQSLACLSEADHTVGFILQ 1462
Db      721 LSNFMEQLKTLRYNPLPEALDPVGVHOBRTLOCLINAVTVQSLACLSEADHTVGFILQ 780
Qy      1463 LSNFMEKMHFFHLPOLMRDIOVNLGYLCOACTSLHSRMTLOHYLQNNKXGDLPSAVQORV 1522
Db      781 LSNFMEKMHFFHLPOLMRDIOV-----AQDGVLESVMLGDRRAVSHMGTPELQDVP 834
Qy      1523 QR 1524
Db      835 ER 836

```

## RESULT 24

```

US-10-719-385-25
; Sequence 25, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1

```

SEQ ID NO 25  
LENGTH: 525  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-719-385-25

Query Match 26.9%; Score 2420; DB 17; Length 525;  
Best Local Similarity 91.6%; Pred. No. 2.3e-210;  
Matches 478; Conservative 13; Mismatches 21; Indels 10; Gaps 1;

QY 1242 LQEVIALFDQTRSLALGSATEDKSMETDSCSRSHRDQDVCVGLHLAKELCEVD 1301  
DB 4 LQEVIALFDQTRSLALGSATEDKSMETDSCSRSHRDQDVCVGLHLAKELCEVD 63  
QY 1302 EDGDSMLQVTRRLPILPTLLTTELEVSLEKONLHFTTEATLLTLTARTQCATAVAGAG 1361  
DB 64 EDGDSMLQVTRRLPILPTLLTTELEVSLEKONLHFTTEATLLTLTARTQCATAVAGAG 123  
QY 1362 ITTOSICPLISVGLSTNGTAQTPSASRKSLLDAPSWPVRLSMSLEOLKTLRYNPLP 1421  
DB 124 ITTOSICPLISVGLSTNGTAQTPSASRKSLLDAPSWPVRLSMSLEOLKTLRYNPLP 183  
QY 1422 EALDFVGHQERTIQCLNAVAVTVQSLACLEADHTVGFILQLSNPKMWHFLPOLKADI 1481  
DB 184 EALDFVGHQERTIQCLNAVAVTVQSLACLEADHTVGFILQLSNPKMWHFLPOLKADI 243  
QY 1482 QVNIGYLCQACTSLHSHKMLQHTLQKNGDGLPSAQAQVQRPSPA-----NSA 1531  
DB 244 QVNIGYLCQACTSLHSHKMLQHTLQKNGDGLPSAQAQVQRPSPA-----NSA 303  
QY 1532 APSSKQPADTESEQAALHTVOYGLIKITLTAALRHFTPDVCOILLDQSLDAEYN 1591  
DB 304 APSSKQPADTESEQAALHTVOYGLIKITLTAALRHFTPDVCOILLDQSLDAEYN 363  
QY 1592 FLFALSTPTPTFDESEVAPSFGLTLATVVALNMLGELDKKEPLTOAVGSTOAGRTTL 1651  
DB 364 FLFALSTPTPTFDESEVAPSFGLTLATVVALNMLGELDKKEPLTOAVGSTOAGRTTL 423  
QY 1652 KSLIMFTWENCFYLLISQAMRYLDDPAVHPDKQKMKOELSELSTLISSTRYFRGAP 1711  
DB 424 KSLIMFTWENCFYLLISQAMRYLDDPAVHPDKQKMKOELSELSTLISSTRYFRGAP 483  
QY 1712 SSPATGVLPSPQGSTSKSPSQEPLIQLVQAFVCHMOR 1753  
DB 484 SSPATGVLPSPQGSTSKSPSQEPLIQLVQAFVCHMOR 525

RESULT 25  
US-10-437-963-135530  
Sequence 135530, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 135530  
LENGTH: 2026  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(2026)

OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_37198C.1.pep  
US-10-437-963-135530

Query Match 4.1%; Score 369.5; DB 16; Length 2026;  
Best Local Similarity 18.1%; Pred. No. 9.9e-22;  
Matches 383; Conservative 327; Mismatches 653; Indels 755; Gaps 98;

QY 25 LGRSALRELSQIEA---ELNKRRLLEGSLYRPPSPSSAEKAKANDVAS----- 73  
DB 45 LDRRAAASPVPDALARIRAHAMLRGVSMPFKPDASGALDABEVVGEHRLAVKP 104  
QY 74 PLKELGRISKFGGLDEQSVQLIQYQEDYRGTRSVKTVLDQESQALILIKIADY 133  
DB 105 ELKAAARLRSKMKMLDEVQYILV-----KTSFTPTALVAD---TESFLVSVQY 154  
QY 134 YEERTCILRCVILMLTYFODERHRYVEYADCVDKREKELVSKYQOFE-----LYKTE 188  
DB 155 YLERQCLLKICIRIFVHND-----CSDSIDAVEEASVLVREEVQRLSIVRDL 206  
QY 189 APWETHGNLMTREQVSRWFOCLREQSMLEIFLYAYFEAPSD---LIVLTMYKE 245  
DB 207 ASAFSVKGG--AEITLIS--WEETLIRINLIFDIPLFF--YDNLSCRNGGLMIMLCIFKD 262  
QY 246 QGFSRQTNRLVD--ETMDPVRDRIGYESA---LIVGMDIESLHCALDD---RRELI 298  
DB 263 MLSSGYDVKFAVSEAKNSF---HYAKQDLITLQTLDFESLIRVDRDEVPSSGYS 318  
QY 299 QFADGLICQDMCLMTFPGDIPHAAPVLLAMALLRHTLNEEESVVRKIGTALQANV 358  
DB 319 TFSVVDLLEMDVEYSKLPFAFAVESGPLILAMAF-----LCVMSLPGSNVNLGP 369  
QY 359 FOYVTRLLQSIAGGANDCTTSYACVCYGLISFVLTSLHHTLGNODIDITACEVLADP 418  
DB 370 VSGFGILRTFIS-----AFVASY----- 388  
QY 419 SLRPLFMCTEPTSGILGILDSVCGMFFHLISPLLQ---LRALVSGSKTAKKVSFLDKS 476  
DB 389 ---EISYQTEBS--LGMILNLTCEVYDGEESLCQFMDKDSFIDGPIR---FVLEEMN 440  
QY 477 FYNELY--KHKPHDVISHEGDTLWRQTPKLLYPLGQGTNLRIPQGTGVQV---LDDRAY 532  
DB 441 GVTLTYAVPRSDTDMVNHQI--EHSPIRIGIEGT--IGSGSIGILYKXLEDVA 495  
QY 533 LVRWEYSYSWTLFTCEIEMLLHAVSTADVIQHCQVRKPIIDLHVXVISTDLSTADCLP 592  
DB 496 LVRWE-----DLCLALHLADKSLAVQASQN---LGYIDKVRIDIAKIFC--- 537  
QY 593 ITSRIYMLQRLT--VISPVVDYIASCNCLTYLAARNAKWTDLRHGTFLPVHAPV 650  
DB 538 --TSIFKYVEDPNNACVMSKTLGMLAEMLSCV-----PYNVVALDQGP--FTYQ-- 584  
QY 651 SSLSQMTSAEGMNAAGVGNLLMNSQPOGEVGTIAPRLTLTYVKGQ-----GSTQSQ 705  
DB 585 ---SGVASDMLBSGALARKMLFATSBDGDCS-----SLTTYLDFAIQLVRLGCAADD 635  
QY 706 GLVPCVMFLKEMLPYHKRYNSH-----GVREQIGCLILEY-- 743  
DB 636 IISFIIPSVQYIWMNHNWKKYSRWKITLKYFDLVKSCIQVKSFSKLGIGIIMETIL 695  
QY 744 ----IHALL--NLCHETDL---HSHRPSIQFLCISGLATBAGQYVINIMGIVPDI 793  
DB 696 YDSSHSVYLHILMSYQLLHSHSGSYCHDKDI-----EDQLVLCGCFDIYV 744  
QY 794 MMAAOPRSDGAEQOGOLIKTV--KLAFSVTNVVRILKPSNVVSPLEQALSOHGAH 851  
DB 745 YMLSNLP-----EREKSKYRLRVGSKSARYYGFVLVKTIA-----FRH 786  
QY 852 GNNILAVLAKY-----IYHKDPLPRLAIQLIKRLATVAPMSVYACLGND--- 897  
DB 787 SCNIIDVLSFSPSIEWRTCIIFH-----LQLAIFIKILS-----QSHVACSNSEDNN 834

```

QY 898 -----AAA-----IRDAFLTRLOS- 911
Db 835 RTSNKPMDTNDTDFLKAIVAAARVFSMLCTAVAKAPOLMENAYFVNSSEIMRLQTS 894
QY 912 -----KEDRIKIMLEFIVAVETOP-----GLIFLFLNLEKQDS----- 949
Db 895 ISCLIDEVDKNEVVAIFNLISARQADAVLWLRVHLPEOSSTRAQADSNSAH 954
QY 950 -DGSKEFSLGMSCLHVALELI-----DSQODRYWCPEPLHRAAIFALHMODRRDS 1002
Db 955 EQSKYFPLNBSGSPRLVEQILGIGSTELMDR-----SPSILSGVDLILKALMESGAQF 1011
QY 1003 AMLV--LRTKPEMENTLSPLEGTL--SPSSET-----SPSILTECAL--INKIICLETIY 1053
Db 1012 IYILEKRSRTFMENLSCCIRAAPASYPIDSVETDEKSLRYCCLETIFEIMSYELFL 1071
QY 1054 VVK-----GSLDQ-----SLDYLTKKSEIEKRFAYMS 1080
Db 1072 QGKLLTETKTSDDPAVSGKEQKPSVADCPSDIYLKMPDSTTMEBLVNHLS-----N 1124
QY 1081 GYVKSIAVHVAETEGSSCT-----SLEYQMLVSA-----WRMLL 1115
Db 1125 GYQNDL-IHRAKAASCRCIILLTLTSGDGTSGLSFSLVKKIQLISSKULTMCGAKRTLL 1183
QY 1116 I-----IAT-----HADIMHLTDSVVRQLFLDVLDTYKALLVPAVSNCLRIGS 1161
Db 1184 CGGPRIASGTNGQVWTHRRHGDRAOSTDA-----NAFGLQGVVTFP-----CI----- 1227
QY 1162 MKCTLLILLRQWKRELSGYDE-----ILGPLEIL----- 1192
Db 1228 --CSSTVITCPSWLQMLIFMDEPNFLHDIMLKLIHAELEGRPTTPEPQELLCFLLEF 1285
QY 1193 -----EGVLQAD-----OQLMEK----- 1205
Db 1286 KVFENHSEQLOKSPPAANGVSLFDVPHIRDELGLMNSDMKYKAVAKMLDIMHKA 1345
QY 1206 -----TKAKVSAFTVLQW-----KEMKVS DIP-----QYSLVLNVCETLOEV 1246
Db 1346 NLMKQVDAKLALCARSLFITFLSVCTGSSYKFGIFGGGISITTTQSANRCAKCSLQSAV 1405
QY 1247 IALPDQTRHS-----LALGSATEBKOS-----MENDDCRS 1277
Db 1406 DSLPEVDNSGVLPPLSGVVELLITRILLDHAKQSKSSRHLYPVIVILMKTSIGASTS 1465
QY 1278 -----RHRDORDG-----VCYLGLHLAKEL 1297
Db 1466 FLFLMPPSPALKQPVKSLVLLSLPEFIKYKQDMKQSDVAINFGLSLSLSPVL 1525
QY 1298 CEVDEBDG-----SWLQVTRRLPILEPILTLTTEVSLRMKQW--LHFTERTL 1341
Db 1526 CKLAESREYFDLAIASMDIILKGFPLSNVWPVPILOKHF-RLQVILLOKQSGALICTQVIL 1584
QY 1342 HLLTLARTOOGATAVAGAGITQSGICPLSLVYOLSTNGTQOTPSASRKSIDAPSWGCV 1401
Db 1585 NFLITMERTKGAALIKLOSANIFAFIKV-LIS--QMSLDDSCLRSLSTQYTDVKIM---- 1637
QY 1402 RLMSLMEOLKLT-----RNFLEPEALDF-----YGVHOERTLOCIINA 1440
Db 1638 GLGLAIVSLNHCMDDDISRNSVANSTISFLSGVPLMSSVLSAQSVAVTHQSKRTILOK 1697
QY 1441 VRTYQSLACLEADHTYVGFIIQLSNFMKEWFFHL-POLMRDIOQVNLGYICQACTSLHSRK 1500
Db 1698 SQT--SISLALSTERN--ITLCLILAKYHPRDQTKKEVDSEL-----RE 1738
QY 1501 MLQHYL-----QMKNGD-----GLPSAVAQRY--QRRP-----SAAS 1530
Db 1739 ITHLLAFISRGSEKTSQPNWNLSFGCPPIIKKEEMKLNBERPLRSKYGFRPAACCTL 1798
QY 1531 AAPSSSKQF-----AADTEASEQO-----ALHTYOYGLKLIL--SKTLAAL 1569
Db 1799 STPEVSGPNNAGLGLVIRKDKNPAQSDSKQRTFEMLAVOGYIRAFILMKRLKLGQAKNAV 1858
QY 1570 R-----HF-----TPDVCOILLDQSLDLAEINFLFALSTPTTPTPSEVAPSPGT- 1613

```

```

Db 1859 FRABELEFLDLAPPELMPEDILHGLQD-----VSVIVTEVLEAVSTALNTE 1907
QY 1614 -----LATVYVALNM 1624
Db 1908 TERVQCLLVILETSLYM 1925

```

## RESULT 26

```

US-10-106-698-5917
Sequence 5917, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:

```

```

APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
FILE REFERENCE: PA005P1
CURRENT APPLICATION NUMBER: US/10/106,698

```

```

PRIORITY FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28

```

```

PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-11-03

```

```

NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 5917
LENGTH: 63

```

```

TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (36)

```

```

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5917

```

```

Query Match 3.4%; Score 305; DB 14; Length 63;
Best local similarity 96.6%; Pred. No. 1,7e-19;
Matches 56; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 953 KEPSLGMWSTLHVALELIDSOQODRYWCPEPLHRAAIFALHMODRRDSAMVLRK 1010
Db 6 QERSLGMWSTLHVALELIDSOQODRYWCPEPLHRAAIFALHMODRRDSAMVLRK 63

```

```

RESULT 27
US-10-437-963-173637
Sequence 173637, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:

```

```

APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua

```

```

APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.

```

```

APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

```

```

FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 173637
LENGTH: 1745

```

```

TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_71655C.1.pep
US-10-437-963-173637

```

```

Query Match 2.7%; Score 245; DB 16; Length 1745;

```

Best Local Similarity 18.8%; Pred. No. 1.6e-11;  
Matches 338; Conservative 279; Mismatches 623; Indels 554; Gaps 88;

```

Qy 270 GYFALLIVGMDI-----ESLHKALDDRREHQAQDGLICQDMCLMTFGDIPH 322
Db 18 GVFFSLVYTFVSDALSTSHKPSLSSSDSFRDHEH-----VMSDNMLTEGFGV-- 71
Qy 323 HAPILAAALRLHNLNPEETSSVAKIGTALQANVPYLRLLQSLASGNDCTTSTAC 382
Db 72 ---VRLAAV--HMLLTQDRSSA-----RDTLTSSRDVDIWAQ 106
Qy 383 M---CVYGLSFVLSTLEHTLGNQODIIDPACVLAADPSLPFLFWG-----TEP 429
Db 107 LEITCRQMSFOGL-----QERIMQTRAYKNDEBDIVMTGTMLHMLMCFLSHP 155
Qy 430 TSGGLIIDSVCWCPHLLSLLOLRALVSGSKTAKV-----YSLDKMSFYNELK 484
Db 156 TSBDKI--KEIKEKTMALSPYGSIRDRHEDPSRTGEIQPTNQPFISLELVREIYQK 213
Qy 485 KPHVISHEDGTW-----RQTPKLYPLGGQNNLRIPQGTQGV--MLDDRAX- 532
Db 214 EPELVHGER--LWTFVISAGEDHTNLTLLVAFGLSLSTLASSEVGAAYELLQKYYR 271
Qy 533 LVRWEYSVSWTLFTC---ELEMILHYVSTADVIQH-----CQRKPIIDVHKVISTD 583
Db 272 SLGMS-----TLRDCLSIYEKKPKESLOSASVWPEPEADQALVSYLAVLQKVENG 325
Qy 584 LSIADCLLPITSR-----IYMLQRLTVVISPPVDVIAVCVNCCLTVLAARNA--KVM 634
Db 326 NT-----TERRKWPFDIEPLFKLSYENVPPF-LKGLARNSITARTIKVSPILKDAIW 376
Qy 635 TDLNHTGLPRVAHPVSSLQWISAEQNAAGYGNL---LANSQPOEYGTIAFLRLI 691
Db 377 SYLQOYD-LPVATPPLGQ-----HNATOIYMRLENEVEWRESYSTISFLMW 426
Qy 692 TTLVKGQLG-STOSQGLVPCWFLKEMLPYHKRKYNSHGRVQEOIGCLIELHAIINL 750
Db 427 NALLAEENINIDKGRFMGIRKFTYEDVFPFQQAIVADPREKELAAVACIEHPFMVLSM 486
Qy 751 ---CHETDLH-----SSHT-----PSIQFL--CICSLAYT----- 775
Db 487 YDIKODDIYAAMNAGSPSTSHASIDROLPVLELKKSCTLSLPHVILSAIPVIKENPDF 546
Qy 776 ----EAGCTVINIMIGVDTIDMWAAQPRSDGAGGCGQLITTYVLAASVTNNVR- 830
Db 547 MSGKVAFNINIVISVGDITL-----INERTTQYTGILLEKTYHLSFEIFLWNER 597
Qy 831 -----LKPPSNVSPLEQALSHQAGNNLAVLAKYIYHKDPALPRALIAKRLA 883
Db 598 DVLVADVRP--LYQPLDVVLAQNHRH---IALLERFYDYLPJOQCSIKIMGILS 650
Qy 884 TVAPMSVYACLGNDAA-AIRDAFLTRLQSKIEDNR-----IKWILPELTVAV-ETQ 933
Db 651 SRIYGLVQLLKADVAKSVIEDYAACLFRRPDPQAIENITKODVGLLQLLIDNICPA 710
Qy 934 PGLIEFLNBYKQSDGSKFSL---GMSCLHAYLELDSQOQDXYWCPPLHRAI 989
Db 711 PNITHTLIRPDV---NGSIRTYVAKPKSHVSCLETTIINDNEKYKTPD--INALHBERGF 764
Qy 990 AFLHMLMDRRDSAMLVLRKPKFWENLTSPFLGTLSPSETSESIETCALIKIICL 1049
Db 765 QLTYELCLD-----PL-----TCGPVNDLST 786
Qy 1050 EIVY-----VVKSGLDQSLDYLTKKFSIERKFAVWSGVYSLAVHVAETGS 1096
Db 787 TKYQFSGHVGITIGVSPLRNNNSLRISM---LHER--AWLLKMLALHVSDDISS 840
Qy 1097 ---SCTSLLEYQMLVAMRMILLIATTHADIMHLLTDSVVRQLFD-----VLDGTKA 1146
Db 841 LYRSEC-----LAILCHTFG--HCAENLRSANLQSSCGSSNLANNGKV 882
Qy 1147 LLLV-----PASVNCRLGSMKCTLLILLRLQKRELGSV-----DEIL 1185

```

```

Db 883 LDLEWVQFRCPDPTSIKYPQLSMRLSES---KIEILRNSATSEFGVYYSERGRLI 939
Qy 1186 GPVLEILEGVQAOQ---QIMEKTKAKVFSAFTV-----IQMKEMKXSDIQY 1231
Db 940 -DLDAFHEKLLQMSQLNPNLQSESEKELKESFQMKWAMRYKNLEBOAQLHMLTGW 998
Qy 1232 SOLVIANVCEQTEBIALFDQTRSL-----ALGSAATEDKOSMETDSS----- 1275
Db 999 SQIV---EIVASRMSILERSHLLPELDASLSAT-----TSPDCSVMAIYLITVA 1048
Qy 1276 ---RSRRHQR-----DGVCVLGLHAKELCEV-----DEGDC-- 1305
Db 1049 LTCMAKRLDERFICPTGADSDAVTCLDISAKQSSNACTSLPFLTVAIIRNSESSELR 1108
Qy 1306 -----SWLOVTRRL-----PIPLTLITTB----- 1325
Db 1109 RROYALLSYFOYCRSTLSDSDVPPVLRFLLEBOGDDDELQKVKLEONEELARSNS 1168
Qy 1326 -----VSLRMKONLHPTFA-----TLHLTLTARTQCATAVAGAITOSICLP 1369
Db 1169 IIRBAQAVIDLVAKDAIHGSEAGKAISFYVLDSLISIDHDKYFLNQIQSGILRS-CLS 1227
Qy 1370 LLSVYQSLSTNGTAQTPSASR-KSLDAP-----SWPGYRLS----- 1404
Db 1228 DVNNY-LSKEASFSESSQRCFTIDAQSLILRLISHYGHKGSQILSMGALHNLSCNL 1286
Qy 1405 MSLMBQLIKTRYNPLPEALDFGVGHQERTLOCIANAVTVOSLACLEBADHTVGTIOLS 1464
Db 1287 MGSQKKANSRLNSVYVERAGEIDKRSLTAPIIRIYTSFTSL-VOSAD----- 1334
Qy 1465 NFMKEMHFTL-----PQMRDIOVNLGYL-----CQAC-----TSLHSHR 1499
Db 1335 -FLBMYEILAYAGTTSARTDCPMLGFDL--CLIVYDADSWIGCKGCPGHAMFTINSL 1391
Qy 1500 KMLQHYLQNK---GDGLPSAVQORVQPPSAASAAPSSKQPAD-----TEASEQQA 1550
Db 1392 VKIRPFEDENLASGDNVKKVMTAIFRLALEAAR---QPSYDRRLKRWMSQATKNQ 1447
Qy 1551 LHTVOYGLIKTL-SKTLAALHFTPD---VCQIILDSGL-----DLAEYFLPALSTT 1600
Db 1448 -KEVRKGINSLVKKKIYVELDFAKOHQPVENIILRESISANIFNERLIMVVISLQKV 1506
Qy 1601 PTPDSEVAPSRGTLTATVNAVNLMLGELDKKEPITQAVGSTAEGRTLSKSLMPTME 1660
Db 1507 WAYENDECSFVQDLFSM---MQFLPSLIDISLNMPOS---SNMIENOKS--ELIVFL- 1557
Qy 1661 NCF-----YLLISQAMRYLRDPAVHPROKQMKQELSELSTLSLSRYFRGA 1710
Db 1558 -CFSLISLYLVLAIRKKMRPQISYDPTTEGGQOPTLQILSDLNSITVAMERVA 1610

RESULT 28
US-10-437-963-154118
Sequence 154118, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 154118
LENGTH: 2905
TYPE: PRT

```

ORGANISM: Oryza sativa

FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_54008C.1.pcp  
US-10-437-963-154118Query Match 2.0%; Score 176.5; DB 16; Length 2905;  
Best Local Similarity 18.8%; Pred. No. 6,8e-05;  
Matches 323; Conservative 249; Mismatches 599; Indels 543; Gaps 86;

12 SPCRSSRELMTLLGRSALRELSQIEALNKRRLLEGLSY-----KPPSSSAEK 64  
40 SFCR-----LLARNP-ANLKPALHPASATWPLMALTKCVIADISGKSPKSA 90  
65 -----VKANKVASPLKELGLRISKEFLGDESGVQLQCLQ 102  
91 RMLRLVACAOADFLRSGSLLLISVAKKLFSHIMEVVDABFL---TEYSILRQLITV 147  
103 EDYGRTRSVKTVLDEROSQALIL--KIADYEE-----RTCLIRCVLHLLTFQ 152  
148 KEYR-----YQMKRTYSSFFVLVYMKKVATGPEKISNQASSKESEFRCLTILHVL 200  
153 DERHPY-----RVAYADCVDKLEKELVSKTRQPEELVKT-----EA 189  
201 NPPGDPIMREESLFTYAKVQELARAPILKLDVILKEIDQMFNTGAGLVWCEA 260  
190 PTWETHGNLMTREOVSRWFVOCLEOSMLLEIFLYAYEFEMASDLDVLTGMEQGF 249  
261 PRDEAAGSL-----RCFOELMDLSATALYLA-KTPRTSHNKKLKT----- 304  
250 SRQTRHLVDFTMDPVDVRIGYFSAIIVEGDIEISLHKCALDREHLOFADQLICOD 309  
305 -----HVTAIMDS-----LLRGLSVMS-----GSLC-- 326  
310 MDCLMTLGTGDLPHANPVLIAALLHNTLNPEETSSVAKIGTATQLVNFO---YLTRL 365  
327 ---CLHKYK-YKVDKSLINMF-----KATCKGLSVLSNNAALF-PDSLLMLIRV 374  
366 LOS-----LASGNDCTSTACMCVYGLSFVLTSLHTL--GNODIDITACEVLADSL 420  
375 LKEPALLMSAKRDP-----PFWLINDKMSILGTYODTNSLIHAL----- 417  
421 PELFWGTEPTSGLGILLDSVCGMPHLL-----SPLLQRLALVSGSTAKK 467  
418 -PLF-----STTALVDSVLRLGEMIMRDVHASFVSEDTWDLQIKQPS--SSTLYF 469  
468 VYSEPLDKMSFY---NELYKXK-----PHDVISHEDGTLNRQT---PKLLYPL-GGQT 513  
470 IACFFSKIGFQGDVNSIFIRKMLLRSTFELVHSGFSLNBOQVLMIPETIFSLCAGFS 529  
514 NLRIFFQGTGVGV-----MLDDBAYLVWEXYSYSWTLFTCEIB--MLHVVSTADVI 563  
530 SPVINSATLSQLFGECKLSDKCKMSHEEELGYSVEITSEINLBPTQAMSEKSRALP 589  
564 QHCQRVKPI-----IDLVHKVISTDL--SIADCLPITSRIYML 601  
590 RHIG--QELLELEFTKGFMSNKKQPEKVDPLNVLVACSLVCMNLHICSL--LSRVF-- 642  
602 QRLTVISPPVDVLAACVNCCLVLAARPAKWTDLRTG-----FLPFWVHP 649  
643 BEKSSFLQVMDLVYVNAIKHIVSVMKHAELSHGLVNLGVSFDTTGSALSSFSFMHP 702  
650 VSS-----LSQMISAEQNNAGVGNLMSSEQPOGEYGTIAFLRLIT 692  
703 LFLSLMRVNNKTSVSVGLTELLDELLELALSELPAQESSLMNPFQGNPKMLPVFVLSSE 762  
693 TL-----VKGQL-----GSTQSGVLV-----CVMFVLEKMLPS 721  
763 DLHPDNRSSFYVMDMDMTGSEVDPVTAGSGSGMWISPRPLECKLELVCIISTFFSVAP- 821  
722 YHKXR--YN--SHGVRQIGCLILELHAILNLCHEPDLHSHRPSLQ---FLCISLAY 774  
822 LHTWEVLNLADKEVNIVGAIL-----DELCTNISASSSSLSASVQDVFLDLNKA- 874

QY 775 TEAGQT-----VINIMIGVDTI--DMWMAAQPRSDGAGGCGOLLITVLAFL 822  
DB 875 TEISFTMPFRIRLINCISNFTYLPVADMT-----GHLL-----SML 914  
QY 823 SVTNVNRKPPSNVSPLEQALSQGHAGNMLIAVL-AXKI-YHKHDPALPRLAIQLLK 880  
DB 915 HDTYRRLVYARKIV-VLQIMEGHSELPHDVSSISGVWVQGSNNPVAR----- 966  
QY 881 RLATVAPMSYVACGNDAAIRDAFLT--RLQSKIEMRKIMWLEFLVAVETPGCLIE 938  
DB 967 EVLAIGQPV-----PIETALTITAHLSLOSEDIEVECVMISAVAIE--PSORE 1016  
QY 939 LFLNLVYKDSGDSKERSLGMWCSLHVLLEIDSGQDRY---WCPPLHRAAIAFLHAL 995  
DB 1017 L-----TYALFDSVSKLNAFASKYDOLIGFLFR----- 1048  
QY 996 WQDRDSAMLVLRK-----PK-FWENLTSPLEGTLSPPSETSEPSILETCALIMK 1045  
DB 1049 WVACEVSLVSEVKEHFGFSDPKNFIEHCSWLSFILRGETADLWLS-----K 1102  
QY 1046 IICLEITYVYKSLDQSLKDTLKKFSIEKPAVMSGVYKSLAVHVAETGSSCTSLLEY- 1104  
DB 1103 ILSQPLSAVYKGYF-----VPIFGLCIARVSGGEKOLA-----ETVLCESLQLG 1149  
QY 1105 -----QWLVSAMRMLIATY--HADIMHLDTSVVRQL-----FLDVLDT 1144  
DB 1150 EISEPERDOLIKGMVSIWGLFSLVSSSSROPELPHSKAVHNSVKTVDGFRDANDGN 1209  
QY 1145 KALLVPAVNCNLRLSGMKCTLLILLRQMKRELSGVEILGPI--TEILEGVLOADQOL 1202  
DB 1210 LAETVVDIKINIFR--SDRVKFLAIHQIADASHPRHRIHRLCAIEVLIDVL----- 1261  
QY 1203 MEKTKAVPFAFTVLOMKEMKVSIDIPOYSQVLNVCTEOEVIALFD---QTRHSIAL 1259  
DB 1262 --GHRVALYSTWFEYIIICVSYIQROPLODO-----CNTLSKLLAFNSCSTETDAL 1314  
QY 1260 GSATEDKDSMETDCC-----SRSRHRDQDGVCVLGLHLAKELCEVEDSDSWLQV 1310  
DB 1315 GKQLODLEPLSPPCCKDIQVFAHLSGSYDSR-----DQTLKF 1353  
QY 1311 TRRLPIIPT--LITTLEVSIRMKONLHFTBATLHLLTLARTQOGATAVAGITOSIC 1367  
DB 1354 VSRAPHLPELFFLSSSVA-----NEASSVL-----ADFSRAGIS---- 1390  
QY 1368 LPLLSVYOL---SYNGAQTSPSASRKLDPASMPGVYRLMSLMEOCLKTRRYFLREAL 1424  
DB 1391 ---DVHOIVFHVNLTETKYPLOPHISKE--DKLPDVGISDDIILVGLLKLACTLSDESA 1445  
QY 1425 DFVGVHOERTLO--CINAVRTQSLACLEADHTV 1457  
DB 1446 EIIDV--TSRTLRGVLSYARGLNALHCLDSIDRSL 1478

RESULT 29  
US-10-128-714-8246  
Sequence 8246, Application US/10128714  
Publication No. US20030119013A1  
GENERAL INFORMATION:  
APPLICANT: Jhang, Bo  
APPLICANT: Hu, Wenqi  
APPLICANT: Tishkoff, Daniel  
APPLICANT: Zamudio, Carlos  
APPLICANT: Erosbkin, Alexey M  
APPLICANT: Lemieux, Sebastien M  
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
FILE REFERENCE: 10182-018-999  
CURRENT APPLICATION NUMBER: US/10/128, 714  
CURRENT FILING DATE: 2002-04-23  
PRIOR APPLICATION NUMBER: US 60/285, 697  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: US 60/287, 066  
PRIOR FILING DATE: 2001-04-27

PRIOR APPLICATION NUMBER: US 60/295,890  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: US 60/303,899  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: US 60/316,362  
PRIOR FILING DATE: 2001-08-31  
NUMBER OF SEQ ID NOS: 8603  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 8246  
LENGTH: 1676  
TYPE: PRT  
ORGANISM: Aspergillus fumigatus  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (1565)  
OTHER INFORMATION: X= any amino acid  
US-10-128-714-8246

Query Match 1.9%; Score 175.5; DB 14; Length 1676;  
Best Local Similarity 17.9%; Pred. No. 3.2e-05;  
Matches 325; Conservative 282; Mismatches 665; Indels 543; Gaps 84;

QY 25 LGRSALBELSQIEALELNKH--WRLLLEGLSYKRP-PPSSSAEKYKANK-----DVA 72  
DB 20 LSDSSFVNIERLRYELETHIEDFKLTD-----KSKNNTSRQAVLSGKITVDVEYSIN 74  
QY 73 SPLKELGIRISKFGIDBEQSVQLQCTLOEDYKGTROSVKTVLODERQSQALLIKADY 132  
DB 75 QDFQOQALQLDALNIDELAAVLEFFAAQER-----AQVLDRLPLIAIMR----- 120  
QY 133 YEEERTCILRCV-LHLTYFODERHPRYEYAVDCVKEKELVSKYRQGFELXYTEAPT 191  
DB 121 FHERHFLBELRLILQESFEVER-----EVTQALMDQVA-----FVLEKXT 163  
QY 192 WETGNMTERQ-----VSRRFV-----QCLBQOS 216  
DB 164 GPLNSSLFARKCKMSMEDIGMILLLEQIQKASIVQVEDRDVWEAIEYQRSLSQOQH 223  
QY 217 MLLEIFLYAYYFE---MABSDLVLTGMFKEQGSGQTRHVLNDEMDFPVDRIQFS 273  
DB 224 ESLALICY--FFKGPTSPEDRLRL-----LNRKRLERFDGL--LHYIP 266  
QY 274 ALI--LVEGMDIESLHKCALDRRELH---QAODGLICQMDCLMTFFGDIPIHAPVLT 328  
DB 267 AMISSFVQHGSPERSN--SYKANSLSHLAVSTKDG-----QWTPIPF---HAAVIA 314  
QY 329 AVALRLHNLNPEBETSVVRKIGTALQUNVFOYLTRLQSLASGNDCTTSTACWCYGL 388  
DB 315 LMLAVSGMDIDGPTSPVQ---GVDPEKAEERTQMFTALDDG----- 356  
QY 389 LSPVLTSLHATLGNQODIITPACVTLADPSLPFLFWG---TEPTSG-----LGII 436  
DB 357 LDFILA---ICSGVNNEMADPARRELVALLIKESAPANLSEPCAGFLKTLNMENTEIF 413  
QY 437 LDSVCGMPHLISPL-----LQTLRALVSGKSTA-----KVVSPFLDMGSPFN 479  
DB 414 VESCIANMPDAVRKLIKSEEDMORLDQITALRDGLSSSLHKGIVREARTLSEFLMIATAAF 473  
QY 480 ELYKHKPDAVISHEDGLM-----RRQTP-----KLPLPL-GGQTNLRIPQSTVGQ 524  
DB 474 ESRDAQEFWADPDGNLYGFLQWASKRGQTVPRVSAFELLCSISGSEHMAAARFLTE 533  
QY 525 VMLDDRVLVWEVSYG--SWTLFTCEIMLHV---STADVIOHCKVRP----- 571  
DB 534 ---EDKFLSSFKSTSMWMSQMFALQLYATRYTEKSPASQAVLRARAKSEPADMSPEBS 590  
QY 572 -----IIDLVKHYVSTDLSTADCLPITSRIYMLQELTTVISPV-----DIVASCV 619  
DB 591 PVMLTCLYRLMGLHCKQSGSIREMQLQHS--FSVSTLTLICSGPITLHRLATVFTTLA 648  
QY 620 NCLTVLAARNPAKWTDLRHGTGFLFVAVPVSSLSQMSABGMNAGYGN-----LIM 672  
DB 649 ALMTDRTSNGNEMWL-----SLDQWISGAMSGPGLGKVPYISNPLVW 692

QY 673 NSEQPOGEYGVTI-----AFRLITTLVKGQUGSTQSGQLVPCWFLKEMLPYHKMRYN 728  
DB 693 HEQDAFOKIGESFQOANHFVLLISLTPTSDESTYHSLD-----PPESIGS-----SYR 743  
QY 729 SHGVREIGCLILBILHAIINLCHETDLHSHSTSLQFLCT-----CSLAYTEAGQTVIN 783  
DB 744 MGLIEPIYIDFVWGQALSRKV-----PDIGERQTLTLTYNCLNPFYITGLRSFNENIIVYLS 798  
QY 784 IMGIGVDTI-----DMVMA--AQPRSDGAEQGGG----- 811  
DB 799 OPTISDSSSLKTSITTYIRLHPARVAEWLFNEVDVKAIFATQOQADASEVANASDSIL 858  
QY 812 -QLIKTVKLA-----FSVTNNYIRLKPPSNVSPLEQALSQ-HGAHGNL--IA 857  
DB 859 IQCLVKSIEVNDMLDLQSTYFHLVRSISQAGSRINVANSSLSFEDSIANNLTITP 918  
QY 858 VLAKYIYHKDPALPRALQILKRLATVAPMSVYAC--IGN-----DAAA 900  
DB 919 ALSLYCGAGHE-QLTVSMALLEBQSSSRKLNKLSPEISWQSNKIIEVLTSTEDVDS 977  
QY 901 IRDAFLTRLQSKIEDMR-----IKWILEFLTYA---VETOPGLIELFLN----- 942  
DB 978 VSRPLVSOQPELRELBEGARSPGYIIRBSILALINCLSMITDRPTIHLFLGFSVGT 1037  
QY 943 -LEVKDSGDSKERSLIGWGLHVLLELDS--QQQDRYWPPLH--RAAIAFLHALMQ 997  
DB 1038 MLDISPN--LFSNGM-SLHRAITGFLQYPDAMBSNIIIPMVHRLRMALVULKILMS 1092  
QY 998 DRRDSAMLVLRK-----PKFW----- 1014  
DB 1093 SKIASATLLEMRSRFLMSFASQPIIGPTPPDGLSVNDEEFWISSTALAEFLFR 1152  
QY 1015 -ENLSPLEGT--SPPSSTSEPSIETCALINKIICLEIYY 1053  
DB 1153 SYLYAVATEIRAAKLAGSQTLQTEILSTLIGNSTTEGETILNPTFEDLPFADLDISH 1212  
QY 1054 VV-----KSLDQS-----LKDTLKFSLEKRFAYWGSYVSLAHVYA 1091  
DB 1213 KIPARLVFLDGIQFVCAKSQADESLVLYNLAVALIQLRKELLSSGQLRPODEQF 1272  
QY 1092 ETEGSSCTSLLEYOMLSAMRMLIIATTHADIMHLDTSVVRQL---FLVDLDGTVAL 1147  
DB 1273 SAEAGLT-----LPIRATNQ-----RLAPRRLYALSMWE-- 1305  
QY 1148 LVLPAVNGCLR--GSMCTLL--TLRQWRBELGSYDEI---LGFTEILBGVLO-- 1197  
DB 1306 -LITMLACSEIEGGRKPTFIHSIQVLPLBAIEBDLBEAIEIARLAETLIGLEFT 1364  
QY 1198 -----ADQOLMEKTXAKVPSAFTV--LQMKEMKVSDIPQYSQVLNVNCELT 1242  
DB 1365 STSTDGARRSGDIIIDEKHLQFOICIRGIFLASGNVQLREFYINISQYIRIAS-ADTV 1423  
QY 1243 QEEVYALFQDTRHSLALGATEDKDSMET--DQCSRSRHRDQRCVCYGLHAKELCEV 1300  
DB 1424 NQNL-----RHHSQOYIKTAGSTLIETDDVAAGEACARVGALEFLN-----LATL 1471  
QY 1301 DEDGDSWL---QVTRRLPILPTLLTLEVLSRKQN-----LHFEATLHLTLTART 1350  
DB 1472 DRETDSTLAEILISQSNVLSLFLDAIRALPHRLKATQANDTPALSYESLSLQRLQOT 1531  
QY 1351 QQCATVAVAGITQSI-----CLPILSYVQLS---TNGRAQRPASARKSLDAPSWGCV 1400  
DB 1532 KNGATQVLTGELFQAVRESRFLPADPDIGIGCKXISLTITDIDNPDAKRYKYYDLG--SV 1589  
QY 1401 YRLSMSLA-----EQLKTLRYNPLPEALDFGVQERTLOCLNATRVQYSLACLEEA 1453  
DB 1590 LRAVLSAVFSGLANEQMOTRAFLAENRQSWGIF--KRFAGIGVGTADHHDLSDL 1647  
QY 1454 DHTVGFILQSLNPMK 1468  
DB 1648 AKCYMTLISATNPLE 1662



```
RESULT 30
US-10-437-963-122168
; Sequence 122168, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 122168
; LENGTH: 2621
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_25121C.1.pep
US-10-437-963-122168

Query Match      1.9%; Score 169.5; DB 16; Length 2621;
Beet Local Similarity 18.3%; Pred. No. 0.00025;
Matches 350; Conservative 273; Mismatches 665; Indels 627; Gaps 86;

Qy 6 KITSVLSRCSSRLM-----TILGSALELSQIEALNKMRLLEGISYKPK 56
Db 481 QILNELACCRMLLEALQKENELELTATLTTEKARKEVEEQRHLCSEKRVLSNLSDEL 540
Qy 57 PPSPAEKVKANKVDASPLKEGLRISKFLGDEOSVOLQO-----YLQED----- 104
Db 541 SLASIKEMENGSNRCADL--EGELRSTK--ENMERTVELASCSNLETLQNDMLLSAN 597
Qy 105 YRGTRDSKYTVLQD-----EROSQALIKIADYYIEERTCIIRCVIHL----- 148
Db 598 SSFEKEALIKLEEDNLCSNKEQGLLDLSTKEB-----LHLSYAKHEHLSHAR 648
Qy 149 ---TYFOD-----ERHPYREYADVCDKLEKELVSKYRQOFEEL----- 184
Db 649 DMEYTFQGLTQLEENITYTSTVDYQITIKELYAKCNVVLGEARNAHONNEACLOSSE 708
Qy 185 -----YKTEAPTWETHGN-----LMTERQVSRWFVQCLREOSMLEIFLYAYAPEMA 232
Db 709 IIVENVERETTSPELIGHDDNORPLVAEND-----SC--NSTALQSL--KGHLKVA 756
Qy 233 PSDLIYLTMRKEGFGSRQTRNRLVDETMDFDRIGYFSALLIVEGMDIESLHKCALD 292
Db 757 KGDRLDQKIVER--ISSRSDGRVLSKLIQSFESK--GNQBDLGMSEBEN-----D 804
Qy 293 DRRELHQAQGLICODMDCLMTFGDIPH-----HAPVLLAMALLHTLNPEETS 343
Db 805 NUKRLTQ-----EMICRLVEKAKAMTSIDAKTEEYVAELCNRIELISYKMSQHEAIEHTA 860
Qy 344 SVVRKIGGTALQI--NVFOYLTRLQLSASGNDCTTSTACMCVYGLSLFVLTSLDLHTLG 402
Db 861 VLVAKMDFAGKLSYKQTIQVLSQVAVHODADNHNAGRLIDQAEILLQNVTERISTLE 920
Qy 403 NQO--DIITFACV-----LADPSLPELFWGTPEPSGIG--IILDSV--CG-----MF 444
Db 921 KERTSLTDVLMVEVTDKLSALSKNALPSDIGSE--GLGSLALSVECAAKLVONTQEKL 977
Qy 445 PHLSPLLOLRLAIVSGKST-----AKKYVSFLDKMSFYNELYKPKHPVI 490
Db 978 EHAQTDNAKNAASLVEELTAHSDVOERSKAHAGVYKMYISLQELLP--NSLGN----- 1029
Qy 491 SHEDGTLMRQTP--KLTYPLGQOTNLRIPOGTVGQVMDLDRAYLVWRWEYSVSWTLFTCE 549
```

```
Db 1030 PDESQVEYNABEPIEALFSQYGDIVERH-----KSLHERQVLYSKNTDLESLSKCE 1083
Qy 550 IEMLIHVVSTADYVQHQRKPKPIIDLVHXYISTDLSTAD-----CLPIRSRIYMLQRL 604
Db 1084 ETEAL-----SSSLTKSNMDFSLNEELKSVSTIEAQDELHGCCLAIAEKVNRSTSHS 1139
Qy 605 TTVISPPVDIASCVNCLTVLAARNPAKVTWDLRHTGFLPFVAPVSSLSQMSIAEGKNA 664
Db 1140 STVLS--SMESSKRNHLLTL-----LPCIEGVASYLE--EFEM-- 1177
Qy 665 GGYGNLLMNSBPQGEYGVTTAFLRLITTLVKQGLSTQSGVPCVMFLKEMLPYHK 724
Db 1178 -----ABE-----IRSKICQESNIGQ--SSSEKMSVSLPVLIKETIYPIF-- 1218
Qy 725 WRVYSHGVRQIGCLLIELHALINLCHETDHSHTPSIQFLCICSLATYEAQYVNI 784
Db 1219 --FDLQGRIDQLSTLNIQL-----ETEPV 1241
Qy 785 MGIGVDITIDWMAAQRPSDGAEGQGLIKTVKLAFSVTNNYIRLKPENVVSPEQA 844
Db 1242 LRQGLTKLDSAL-----ETSAEILQ-----KVFLEQSEKLSYKEX 1280
Qy 845 LSOHGAHANNLIA--VLAKYIYHKDPALEPLALQILKR--LATVAPMSVYA----- 892
Db 1281 LSIIVAKGGLIVQRDSLKQTLLEK--SGELEKLAHELQSKSLILEAKIKSYADADRI 1339
Qy 893 -----CLGNDAALIRDAF-----LTLQSLKTEMRIT-----KVNILFELT 927
Db 1340 EAESELSYRNSXTALRDSFLQDSVYLQRIEVELDLDPENPHFDIVKELLSKMA 1399
Qy 928 VAVE--TOPGLIELEPLNEVKDSDGSKFSLGMSCHAVLELIDSOQOD----- 976
Db 1400 VGASFYTP-----DGKQSSVDGNSGCAIIDSINDQNSNSNGAREIK 1444
Qy 977 -RYWCPPLLRALAFIALHLMQDRDSAMLVLRTKPFWENLTSPLFTGLSPSEETSEPS 1035
Db 1445 IKY--DELHR-----RFYE-----LAENHNMLEQS 1467
Qy 1036 ILTECALIMKICIEIYVYVKGSLDQSLKDTLKKFSTIEKRPAYSGVYKSLAVAVATEG 1095
Db 1468 LVERNNLIQK-----NEEVLQOI--SIPOQRLREPDEDIA--WLQ----- 1504
Qy 1096 SSCSTILEYQMLVSAMMLLIATTHADIMHLDTSVVRROLFDLVDLGTALLVLPASVN 1155
Db 1505 ---NRLLVEVHERNA-----LHKIEHLEBS--SEMILSDLESBK----- 1540
Qy 1156 CLRLGSKKCTLLILRLQMRRELGASVDEILGPTLEILGVLQADQOLMEKTKAVPSAFI 1215
Db 1541 --RISELSAEIVAV-----KAEKEFFSQSL 1563
Qy 1216 TVLQMKEMKVDIPQYSQVLAVNCEITLOEEVIALFDQTRSLALGSAITBDSMETDGS 1275
Db 1564 EKRFDFGLJSE-----KAVODEVR--DWLRKDLA----- 1592
Qy 1276 RSRHRDQDQCVVLGLIAKELCEVEDG-----DSWLOVTRRLPIPLTLLEVSIRMK 1331
Db 1593 -----BLQEKLAKEBTESKLYHDMEMEIKKMDLVADALQD--DSNTEIP 1635
Qy 1332 QNHTEATLHLTLTARTOOG--ATAVAGAGITQISCLPLISYQSLSTNGTAQTPSARK 1390
Db 1636 SGAGVGAIVAVCLGSLIRLLIDGYKTHLSBSTVSSASMETLSETKISKIDASTSERGMEK 1695
Qy 1391 SLDAPSWPQVY--RLSMSLEQL-----LTIKRNPLPEALDPFYGVNOER 1433
Db 1696 EMALNTLSGELEHRTNSLALVEOQDRBAVERKQTLTLETLRQOI--DQIQGDGABQMN 1753
Qy 1434 TLQCLNAVTRYQSLACLEADHTVGFILQSNFKEMHFLPLQMRDIOVNLGYLCOACT 1493
Db 1754 RYQSL-----MUELESMTKQ-----RDLQERTLQGEBOQKCT 1784
Qy 1494 SLHSRMLQHYLQNKNG-----DGLPSAVAQ-----RVORPPASAASAPSSSKOPA 1540
```



Db 1785 SL---REKLANVAVRKGKGLVQHRDLSLKOTMEENMTMEIKLAKVERKQIHESL--ESERSSL 1839  
Qy 1541 ADTASEQOAH-TVQVGLKLKLSKTLAALNHPDPVQOILLDSGLAEVNFALSF 1539  
Db 1840 MGRLEAEKSHDHTQY-----LSRLINSL--TVDIGERPTDPTITVENFSKCLDQ 1892  
Qy 1600 TPTDSEVAPSEFGLTATVNV-----ALNMLGELDKKKEPLQAVGLSTQAGTTLKSL 1655  
Db 1893 NEVKSQKATEL--LLAELEVEHERADYLODELVKAELASEFRKONSVESAR----- 1944  
Qy 1656 MFTMENCYLLISQAMRYLRDPAVHPRDQRMKQ-----ELSELSTLLSL 1702  
Db 1945 -----ADAVRHL-ERIMHMQSKTRKQIDHLMELNSTSSQLREIFSEL 1986  
  
RESULT 31  
US-10-437-963-189782  
Sequence 189782, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Mu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(51321)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 189782  
LENGTH: 2122  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1) (2122)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PRT\_MRT4530\_86258C.1.dep  
US-10-437-963-189782  
  
Query Match 1.8%; Score 166; DB 16; Length 2122;  
Best Local Similarity 19.3%; Pred. No. 0.00036;  
Matches 344; Conservative 245; Mismatches 585; Indels 612; Gaps 87;  
  
Qy 166 VDKLEKELVSKYRQOFEELYTEAPTWETHGNLMTROVSRWFVQCLREOSMLLETIFY 225  
Db 739 IDKLEKCTELVSKYF-----NNDTEALNARILIN--DGNLTLLDDLP- 780  
Qy 226 YAYEMASDILLVLTMRKEGFGSRQTRNL-----VDETMDPFDRIGYSALILVGM 281  
Db 781 --FTTSGKITF-----OKHLYIMVNCIRAPFOISKYFVDEGFSAGV 823  
Qy 282 DIESLHKALDRRELHQFADGLICQ--DMDCMLTFGDIPIHNAVYLLAMALLHTLN 338  
Db 824 RVESLMLMA-----SICSLILGFPCLMP-----LAHE-N 852  
Qy 339 PEETSVVRKIGTAIQANFOYLTRLQSLASGNDCTTSTACMC--VYGLSLFVLTSL 396  
Db 853 KDVISSALKICIEGLSL--VMQ--RLSASLSRNGNG-SKLPRCMLSPFGVF----- 898  
Qy 397 ELHTLNGOODIIDTRACEVLADPSLELFWGTEPTSGGLIIDSVCWMPHILSPLOLR 456  
Db 899 -LGSIVNOKTMSDTRFL-----PAYISLSPSQ 928  
Qy 457 ALVSGSKTAKKYVSLDKMSFYNELYKRPKPHDVISHEDGTLMRQTPKLYPLGGQTNLR 516  
Db 929 DLWMPEN-----LHERFDQSTYDAIILH-----FILRSKMLSSYGFMTLS 969

Qy 517 IPQSTGVQVMD-----DRAYIVRWEYSYSSWTLFCEIEMLLHVSTADYIQ 564  
Db 970 ILKG-VOSILFEDVEDVSKLLFDLLR-----NKQSG-----CESR--QIMSTHEIQI 1015  
Qy 565 HCQRVKEPIIIDLVHKV-ISTDLSIADCLPITSRIYMLLQRLTYVISPVDVIAVCVNCIT 623  
Db 1016 LCILLVWFVSNSANVSSERS--EALLKYL-RIDVGAQDDPVVMP-----CVT 1062  
Qy 624 VLAARNPAKWTDLRHGTPLFVAVHPVSSLSQMIISAGMNAAGYGNILNASEQOGEYGV 683  
Db 1063 ALQAVQP--VEFDLKTIDQEKVAFSLISMERTEITEIRNARDAIILIRINVA-----ST 1115  
Qy 684 TIAFLRLITTLVKQOLGOSTSQ-----GLVPC-----VMPVLKE 717  
Db 1116 AVKFEIELIAOGDKKMSKRIKREDLNHDIKKNPDDLFGVKPAPASVLSLDVLFKGD 1175  
Qy 718 MLPYVHKWRVNSHGVREQIGCL--ILELHAILN-----LGHETDLH--SSHTPSLQ 765  
Db 1176 VI-----QRTCLLOPLFQLSKLSLSDQILGIVQYKNGHASPENPDL 1220  
Qy 766 FLCISLAYTEAGQTVININGIVDTIDWMAAOPRSDGAEQOGOLITVKALESV- 824  
Db 1221 NFMV-----BAQGLVLLIKDITDTLQ-----SGHODELFNCRDINLLINCIOAKDQ 1269  
Qy 825 TNNVYIRLKPSPNVVSPLEQALSQHGAGNNILAVAKYIKHKDPALPRLAIQILKRLAT 884  
Db 1270 TRN-----HGFSLIASLAK-----APQVSSIESIDL- 1296  
Qy 885 VAPSVYACLDGDAARDAIRDAFLRLQSGKEDMRISKVIMLEFLT----- 927  
Db 1297 --FVAID--AKQOLFIRKLVNDVEHRRLITLWYLLRTIGERKCLSTVMCLHS 1347  
Qy 928 -----VAVETOPGLIELPLINEVKDQSGSKESLGMW-SCIHAVLELIDS 972  
Db 1348 LVGRISHSPEHQALSLRAMPQEWEGYLAANI-----TNGYSYKLMFHCHSKLOET-- 1399  
Qy 973 QQDRYRCRPLHHAALAFALHMODRDSMLVLRKPKRWENLTSLPGLTSPSPSTS 1032  
Db 1400 RVHEKQYILPMLH-LAOFILFKLD-----TELIFPL--DSEEA 1436  
Qy 1033 EPSILETCALIMKTIETIYV-----VKGSLDQSLDKTLKKEISERFAYW--SGYV 1083  
Db 1437 ANSIOGSLGELMKVLCVAANVRDKTGISGDALKELRDSAN--TILKVIAGMCASTYF 1494  
Qy 1084 KSLAVHVAETEGSSCTSLLEYQMLVSAMRMILIAITTHADIMHLDTSVVRQPLDVIDG 1143  
Db 1495 KGI-----SRMLBH-----SRSVKRRKAL----- 1513  
Qy 1144 TKALLVPASVNCARLSMKCTLLILLRQWK-RELGSVDEILGPTLEIGVQADQOL 1202  
Db 1514 --GILCETAKNS-----LLOKQKRAKLHST--PAT--ALQVD-- 1548  
Qy 1203 MEKTKAKVFSAF-ITVLQMKEMKVSDIPQYSQLVLANCETLOEBVIALFDQTRHSIALGS 1261  
Db 1549 --KSAPCFSELCKYKILEVDREVDSVSVAIAAISLETIAKYPS--DMPATRKCLAK 1604  
Qy 1262 ATEDKSMETDCCSRSHRRDQDGVCVLGLHAYELCEVEDGDSWLTQYTR----- 1313  
Db 1605 ITNHNSGDVAVTSRSRY-TVGSILINVG--SKALPOLPLIMQMLQVSHQVFCPSGK 1660  
Qy 1314 -----LPILFTLLTYLEVSILRMQNLHFTATYLLHLLTL-----AR 1349  
Db 1661 YAHSTKTDAVLSNOAIPILISVLTVEVI--VKKLEFVNPVLEEIIDLVLVHPECASR 1718  
Qy 1350 TOGATAVAG--AGTOSICL-----PLASVYOLSTNGTAQTPSARSKSIDAPSPGV 1401  
Db 1719 NDEKLDKAKADVRLKLLIDKPVVRLMLSPILNVY--NGAIKCEBA----- 1760  
Qy 1402 RLSMSLMEQLIKTLRYNFLPEALDFGV--HOERTLOCLNA-----VTVOSLACLE 1451  
Db 1761 --SLIAFEMLSLST-----LVGAMDRLANGVHTKYVEHCIVALDLRQHDLSLKNAIAVE 1813

QY 1452 EADHTVGTILQSNFMKEWHFHLPOLMRDIOVNLGYLCOACTSLHSRMLQHYLQN--- 1508  
 DB 1814 QS--IIHATITMLTENTFR-PLFLRTLEMAESVDRSTSKSMRAIYFKLVNSLA 1870  
 QY 1509 -----KNGDGLPS-----AVAQRVQRPSAASAASSSKOPADTE 1544  
 DB 1871 EKRRLGLVLPISVRNMPGWSLFTFYFKYLEGVSQVLSBEDDALISSKOKKKALJEDAP 1930  
 QY 1545 ASEQQAHTVQ-YGLKILSKTLAALRHPTPVQCLLDOSLDLAENYPLRSTTTF 1603  
 DB 1931 VEQKKTSGPKLMNRALVLSLH--KCPLYNDKILDS-----NFOALLKPIVSQF 1982  
 QY 1604 DSE-----VAPSGTILATVNVVALNML---GELDKKEPLTOAVGLSTOAGETRL 1651  
 DB 1993 VIEPEHFESVPEASVDEVDLVLCLGQMAVTRSDVLMKPLNHEVGTHTYASS--- 2039  
 QY 1652 KSLMFTMENCYLLISQAMRYLRDPAYHP-----RDKQMKQELSSSLSTLSLSRYF 1706  
 DB 2040 -----FFLDNGQVLMKTRSDKYRPKMLGLKVRHNVQQLKEEYVVL----- 2081  
 QY 1707 RGAPSPATGVLPSPQGSTSLSKASPSQEPILQVQAFVRHMQ 1752  
 DB 2082 -----PET--IP-----FLAELLEDVELPVKTLAQEIVKEME 2111  
 RESULT 32  
 US-10-128-714-3246  
 ; Sequence 3246, Application US/10128714  
 ; Publication No. US20030119013A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jiang, Bo  
 ; APPLICANT: Hu, Wenqi  
 ; APPLICANT: Tishkoff, Daniel  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Eroshkin, Alexey M  
 ; APPLICANT: Lemieux, Sebastien M  
 ; TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and  
 ; TITLE OF INVENTION: Methods of Use  
 ; FILE REFERENCE: 10182-018-999  
 ; CURRENT FILING DATE: US/10/128, 714  
 ; PRIOR FILING DATE: 2002-04-23  
 ; PRIOR APPLICATION NUMBER: US 60/285,697  
 ; PRIOR FILING DATE: 2001-04-23  
 ; PRIOR APPLICATION NUMBER: US 60/287,066  
 ; PRIOR FILING DATE: 2001-04-27  
 ; PRIOR APPLICATION NUMBER: US 60/295,890  
 ; PRIOR FILING DATE: 2001-06-05  
 ; PRIOR APPLICATION NUMBER: US 60/303,899  
 ; PRIOR FILING DATE: 2001-07-09  
 ; PRIOR APPLICATION NUMBER: US 60/316,362  
 ; PRIOR FILING DATE: 2001-08-31  
 ; NUMBER OF SEQ ID NOS: 8603  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 3246  
 ; LENGTH: 1545  
 ; TYPE: PRN  
 ; ORGANISM: *Aspergillus fumigatus*  
 US-10-128-714-3246  
 Query Match 1.8%; Score 165; DB 14; Length 1545;  
 Best Local Similarity 17.8%; Pred. No. 0.00025;  
 Matches 304; Conservative 259; Mismatches 628; Indels 516; Gaps 76;  
 QY 96 LLOC--YLOEDYGRDTSVKTVLQDERQSQALIKIADYYEERTCILR-----C- 143  
 DB 10 LLECIARLIGBSFEVEREVTOALMOD-----MVAFVVEIKTGPRLNRSLSFARKCM 59  
 QY 144 -----VHLLTYPQDEHRPYRYEAYDCVDLKELEKLVSKYKQGEEL-----YKTEA 189  
 DB 60 KSMEDISGMLLLLEBOIQKASIVGQVEDRDWEAIEYOR--SLIQOQHESLGAILCYFPFG 118  
 QY 190 PTWETHGNLMTROVSRWFVQCLREOSMLLEIFLYAYAFEMAPSDLLVLTQMKEGQFG 249

DB 119 P-----YTPSPEDLAVLNRLKLER-FDGLLHY-----IPAMISFVQHGSP 160  
 QY 250 SKQT-----NRHU--VDEIWDPFVDRIGYFSALIL-----VEGMDIE 284  
 DB 161 ERSNSYKARSILHIAVSTKDGQWTIIPTEAAVIALMLAVYSGMDIDGPTSPVQGVDFE 220  
 QY 285 SLHKCALDDREHLQFQAD--GL-----ICQDMCLMTGTDIPHHAPVLLAMLAHTL 337  
 DB 221 K-----EABERTQMWTALDDGGLDFLALICSG-----V 249  
 QY 338 NPEETSVVRKIGGTAIQLVNFOYTLRLQSLASGNDCTTSPACMCVYGLSFTLSLE 397  
 DB 250 NNEEMADPAIR-----ELVALLIKESAAPANLESP--CAGFLKTLIMENLE 293  
 QY 398 LHTIGNOODIITRACEVLADPSLPELFWGTPEPTSGLGI IIDSVCMPHLLSPLOLLRA 457  
 DB 294 IPVESCIANMPDAVRKIKSEEDMR-----LDQITALADGSS--SLHRG 336  
 QY 458 LVSGKSTAKKYVSFLDKMSFYNELYKHKPHDVISHEGTLW-----RRQTP----- 503  
 DB 337 LVEART--HLESTLMIATAFESRPAOEFPAADPGNLYGFLQMSKQVTPRVSAPFC 393  
 QY 504 KLYPL-GGQTNLRIPQGTVGQVMDRAVLVEMEYSY-SWTLFTCEIMLHVY----- 557  
 DB 394 ELICSLISGGEENAAAHRLFTE--EDKFLSKFKSTSNMWSQMFALQLVATRYTEKP 450  
 QY 558 STDVIVQHQRVPR-----IIDLVHXYSTDLSTADCLLPITSRYMLQR 603  
 DB 451 SASQAVLRARKSEPADNSPEPSVMTLCYRLMGLHGLKSGSIREWMLQHP--FSVAVST 508  
 QY 604 LTTVISPPV-----DVIASCVNCLTVLAARNPAKWTDRHRTGFLPFVAPVSSLSQMI 657  
 DB 509 LTLTCSGPIPTH.RATVFTTLAALMTDRSHNGENML-----SLDQWI 552  
 QY 658 SAEGBANAGYGN-----LIANSBQPGEGYVTI--APFLITTLVKQGLSTQSG 706  
 DB 553 SGGAMSGPGKGVPIVSNPLVWHEQOAFQKIGSPQOANFVVLIVLSLTPSDSTEYHL 612  
 QY 707 LVPCVMEVLKEMLPYHKMYNSHGVREOIGCILELHAILMLCHETDLSHTSLOF 766  
 DB 613 SLF-----PESIGS--SYRMPGIEPIYIDFPMQALSKV--PDIGERQTRLITY 658  
 QY 767 LCT-----CSLAYTEAGQVYINIMIGVDTI-----DMW 796  
 DB 659 NCLNFVITCLRSFNENIVTVLSQPTISSDSLSLTYYIRLHPARVAEMLENDVIK 718  
 QY 797 A--AQRSDAEOGOG-----QLIKYKLA-----FSTYNNVIRLKPFSNVV 838  
 DB 719 AIFATVQODASEVANAASDSILIQCLVKSIEVNDMLDQSTYFHVIRPSIKSQAGGSRI 778  
 QY 839 SPLEOALSQ-HGAGNNL--IAVLAKYIYKHDPALPRLAIQLKRLATVAPMSVYAC-- 893  
 DB 779 NVANSSLSFEDSILANNLTIPALSLYCAGHE-QLTVTSMALLEKMSRKLKXLSPE 837  
 QY 894 LGN-----DAAIRDAFLTRLOSKIEDMR-----IKWMLLEPLTVVA 929  
 DB 838 ISNWOSNKIWEVLSTEVEVDVSVRPLVSQOMPELRELEIGARSPGYIIRRESLALINSC 897  
 QY 930 ---VETOPGLIEFLN-----LEVKDSGDSKESRLGMWGLHVALELDS--QQQDR 977  
 DB 898 LSWITRPFTHLFLFGSSVGTMLDISPNS--LPSNGM-SLHHAITIGFLOSYPDAMEG 952  
 QY 978 YWCPPLH--RAAIAELHALMODRDSAMLVLRTK----- 1010  
 DB 953 NILPMVWHLRMLALEVLKELMSKIASAVTLLMEMRSRFLMSFASQPIIGPDPDGLS 1012  
 QY 1011 ---PKFW-----ENLTSPLGTL--SPSET 1031  
 DB 1013 VNDERFWISESTALAEFLFRSYLYAVATERRAARKLSQTLQTEIILSLGNSSTET 1072  
 QY 1032 SEPSILETALIMKIKLEIYVV-----GSLDLOS-----LKQTLKK 1069  
 DB 1073 GETIILNPTVFDLFPDADDLDSHKIPAPRLVFLDGLGFVCAKSAQADESLVTLNLEAVEEL 1132

```

Oy 1070 PSIERKPAWGSVYSLAHVAETESSGCTSLLEYQMLVSAKMMLIATTHADIMHLD 1129
Db 1133 IQVRKEELLSSGQLRPQDEQPSAEAGLT-----LFRATNOA----- 1171
Oy 1130 SVVRROL-----FLVDLDGTKALLVLPASVNCRL--GSMKCTLL--ILLRQMKELGSVD 1182
Db 1172 ---RQIAFNRLALRWSWE---LITTMALCSHIEGGRKPTFLHSGVLPLKEAAIEE 1224
Oy 1183 EI-----LGPTEIEGVLIQ-----ADQLMKTKAVSFAFITV--LQM 1220
Db 1225 DLPEAIEIARLAETLIGLEFSTSTDGARRSGDIIIDKXHOFOICIRGIFLASGVOL 1284
Oy 1221 KEMKASDIPOYSQOLVNVCEITLOEVIYALFDQRRHSIALGSATEDKDSMET--DDCSRSR 1278
Db 1285 RETFINIGSQIYSRIAS-ADTVNQNL-----RHISQYIKTRAGSTLIETTCDDAVAGQ 1336
Oy 1279 HRDQDGCVCVLGLHLAKELCEVDEGDSWL---QVTRRLPIPLTLLTLEVSIRMKON- 1333
Db 1337 EACRSVALFTNL-----LATDRETDSTLAEILISQSNMYSLFLDAIRALPHEIKNTQAN 1391
Oy 1334 ----LHFEATLHLTLTLARITQAGTAVAGAGITOSICPLISVYGLSTNGTAQTSPAS 1388
Db 1392 DTPALISYIESLISLQRLCQTKNGATQVLTGTFQVRSRSLFPAADPDGLDIDNDAL 1451
Oy 1389 RKSLDAPSPWPGVYRLSMSIM-----EQLKTLRYNLFPEALDPVGVOERTLQCLNAV 1441
Db 1452 RKYVDLIG--SVLRVLIISAVFSRGLHNEQIMQOTRAFLAENRQSMVGIF--KRPAKIGV 1507
Oy 1442 RTVQSLACLEADHYGTFLQLSNFMK 1468
Db 1508 GTADHHDALSDLAKCYMTLISATNFILE 1534

RESULT 33
US-10-437-963-114113
Sequence 114113, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437, 963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 114113
LENGTH: 2462
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_17836C.1 pep
US-10-437-963-114113

```

```

Query Match 1.8%; Score 160.5; DB 16; Length 2462;
Beet Local Similarity 19.1%; Pred. No. 0.0015;
Matches 369; Conservative 259; Mismatches 609; Indels 695; Gaps 106;

Cy 25 LGRSLARLELSQIEA-----ELNKHWRRLLEGLSYKKPPSPSAEKVKYKANK 69
Db 810 LGVEASNRSLSHVIPFLCDATISVGNLKYKQGEHIRKLSIKKNQGEGRSps----- 860

Cy 70 DVASLT-----KELGIRI-----SKFLGDBEGSVQLQC-----YL-----QEDIRGTRDSYKT 114
Db 861 --FSPILVCVLQKICRLRLDSEASVAKHEKSTISLYCNCNTVYLIMQSOVDVLLPDLIVGT 918

```

QY		115	VLOBERSOQLILKADYYEERFCI-----LRCVHLLTYFODEBRHYREAVDCVDKL	169
Dd		919	IINER-----LGRFSSEINSRICEAEMRWPLMLTLNRSIDP-----	956
QY		170	EKEIVSKYROOFEEUYTEAPFTWETHNGIMTEROVSRMFVCLBROSMLT---EIIFLY	225
Dd		957	-----QXSSSLFSTLEHSSEVGNLSLC--SVTRIEMSNOQPNNLDVAITSLY	1005
QY		226	YAFEMAPSDDLTV-----LYTMFKEOGFGSRQTNRHVLDETMDPFVDRI-----	269
Dd		1006	SVI--CAPRPDVISFPRLTHLVKTH-FPSNL-----PFLSSVLPLOHDYLAKV	1051
QY		270	-----GYFSALLVEG-MDBESLH-----KCHL-----	291
Dd		1052	ASYCEPDMEFSSLROIKNMJDVDSVINVEDKMKHSGTCSESAIVISTFLNTPFCALLBSVL	1111
QY		292	-----DDRELHOFAODGLICQDMDCMLTFGDIP-----HHAPVLLAAALRHNTN-	338
Dd		1112	SLASAPBEIRKAPHLODELTV---HLLOAKISESPSELTIPLRVLFMS--HHLLSS	1166
QY		339	-PEETSSVVRKIGGTAYOL--NVFOYLTRLLQSILASGGNDCTTSACVCYGLLSFY--	392
Dd		1166	YTVKCSDIALOCCGVCFELDISIFERI-RVLTADTANSKSVAIFYPVCLNGIYESVOH	1224
QY		393	-LTSIELH-TLGNODIIDTACEV-----LAPSPLELFE-----	424
Dd		1225	P I G I S L C S I S N Q D L A D G S E V Y K E D F A S K E K H L A D S P V L N L T S N U G V L A G S	1286
QY		425	MGTBPTSGLAGIILDSVCGMPHLSPLLOLRALVSGKSTAKKYVSFLDKOSFYNELYKH	484
Dd		1285	FGANYSNNDOSLASLFGPPKLLBERILLRKEPFELCMEGANGFLPNVTMYRTLAKF	1344
QY		485	KPHDIVSHEDGTLWRROTPKLLLYPGGOTNLRIPOGTVOYMLDRAYLVWEYS---Y	540
Dd		1345	-----VSPVRLE--LANMMESTFEFG	1364
QY		541	SSWT-----LFTCEI--EMLIHVYSTAD-----VIOHC-----QRVPI	572
Dd		1365	SSSSPAVAPAACCITYTDIAFEMUYDLOQIDORSGPCRMGLEIHNCIDIATIQOYVNI	1424
QY		573	IDIVHKVISTPOLSIAD-CLIPITSRI-----YMLOLR--TTVISPPDVVAS	617
Dd		1425	I--LHPATKLNLEPRADCLLMOLIRIHHTEIISAVRNDYILAHMLSTMVANTPIDIIHH	1488
QY		618	CVNCLITVLAARNPAKWMTDLRHGTGLPFVAHPVBSLSOMISAEGNANAGYGNLTMNSOP	677
Dd		1483	CM-----FPSIKYAKAIQULLGAGNPAMHMLPCKLTLTD-----	1515
QY		678	QGEVGVTIAPLRILTTLVKLGQGSTOGSGLVPCVMFVLKEMLPSTYHKR-RYNSHCVRBOI	736
Dd		1516	-----LIKDOTSVMO-VYGSDSNASWTHEDSFIL--LLPALSYIEHHSQNRQ--	1561
QY		737	GCLILELIALINLCHETDLSHSTPBLQFCISLATGAAGQYVINIMGVDPDIWMV	796
Dd		1562	GSUNNE-----IAPEKESKR VNNAKKFTSIL-----VRLDXKF	1596
QY		797	AAOPRSDAGEGOGOLIKITVTKLAFSYTNNAVIRLKPSNVVSPLEQA-----LSQHGAH	851
Dd		1597	RNFPRBDIILSSPEBLNVTC-LEVALKRIVELS--SEVOHNLNQKPIPLNQ-----	1645
QY		852	GNNLIJVLAKYIYHKHDPALPRALIQILKRIATYAP-----MSVYA	892
Dd		1650	-----LIRSSILHRFSDP--VWKAIRCIIVLSEBKFPADLELILGHSHVFSTYT	1700
QY		893	CLG-----NDAAAIRDAFLTRLOGSKIEDM--RIKMILDELTAV	930
Dd		1701	CSGVSECPACNPTGGLIOPAPSLIKSVDSAPAR--ENKFPDCCIPERKKVBIIRLLRVLY	1758
QY		931	ETO-----PGLI-----EL-PLNLEVKOSDASKESFSLGMQCLHAVELILISO-----QQ	975
Dd		1759	DIKSOHQHSSLLDSRELRGIFLLSVYGAUTSETDLEI-----LHMNIESSECAKINDV	1813
QY		976	DRYWCPLLHRAAIAFLHALMOD-----RRDSAMLVLRTPKPKWENUJSPLRGTSPP	1028

```

Db 1814 DMLW-----GKAAVFRRELLEFASADTHOMENAEISDRRSLFREMI--PI-----D 1860
Qy 1029 SETSEPSILETC-----ALIMKIICL-----EYIVVKSGLDQ-SLKOT--LKRISIE 1073
Db 1861 SKLCVMVTLQPCYRRSSRASIFSLQLOQDNFGDIFKATOSQMAVRYDVFILHSIH 1920
Qy 1074 KRFAVMSGYKVS-----LAVHVA-----ETGSSCTSLLEYMLVASW 1111
Db 1921 ---TLNAGYIEPVERSRGLAIIITVLSISPEDELRKIGYSLGTFPKKSLASOKSKETW 1977
Qy 1112 RMLLIATTHADIMELTD-----SVRRQDELVDLDT-----KALLVPASVN--C 1156
Db 1978 QLOLLFTYLONGISEWQORIPSVIAIFAAEASLTLDSSHOTAPATISKFLMHSASVNLQL 2037
Qy 1157 LRDSMKCTLLILLRQKRELGSVDLGLPTEILBEGVLOADQOLMEKTAKPSAFIT 1216
Db 2038 LYASGN-----LADDAKIYKR--GGV-----LEIALSYGSSAVSDSETKLLT--LQ 2079
Qy 1217 VLQKEMKVSIDIPOY-----SOLVNVCELTQOEVIATPDOTRSLA--LGSATEDKDSM 1269
Db 2080 VLK-KCVTLFVLAHHLTKDSGLIMISSVISHYEG-LDSVKNSTSYIVIGSALB----- 2132
Qy 1270 ETDSCSRHRDQBDGVCVGLGLHLAKELCEVDEDEGDSWLVQYTR----- 1312
Db 2133 VVNDLTSR-----LIAE-----WLOETALBQLSRISKYLVYVED 2168
Qy 1313 ---RFLIPLTLLTLEVSLR--MKQNL--HFTETHTALLLTARTOGATNAVAG 1361
Db 2169 MKLKGNVPLLTSLVNLVASTMRLSMKRKIIQPHFS--LSLHGIKLCTI-----GG 2219
Qy 1362 ITGSICTPL--LSVOLSTNGTAQTPSARSKSLDAP--SW--GQVRLSMLMEOLIKT 1414
Db 2220 ISRIEIVKLAQGLDVIYLMNGPLPYLSMDKSMRTATVSMATSNIFNL--CDEGRSVLK- 2277
Qy 1415 LRYNPLPBLADLFVGVHDE-----RTLOCLNAVRTVQSLACLEADHTVGFILQSN 1465
Db 2278 ---MP-----HEEPLKNECLLSKMLRWLASIILKICIS-----LOHYIQ 1507
Qy 1466 FMKEMHPLPOLMEDIQNLGYLOACTSLHSRKM-----LOHYIQ 1507
Db 2311 ---HEKGDPLTDAN--NFSL--ESFLNTYDEKVEVTSVSHSADDEALAIILLYLQHLK 2363
Qy 1508 NKNGDGLPSAVA-----QVORPPSAASAAPSSSKOPADTEASEQALHTVQ---X 1556
Db 2364 -MNRDPLPSVVAALCLILDRSNKQGSNHVYSPRHYQPMKDPAMHNRBAHELEBQACQ 2422
Qy 1557 GLIKILSKITLAA 1568
Db 2423 SLVWFNSPSFA 2434

```

RESULT 34  
US-10-408-765A-354

```

; Sequence 354, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Marnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408.765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 354
; LENGTH: 3859

```

TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-408-765A-354

Query Match 1.7%; Score 157, DB 16, Length 3859;  
Best Local Similarity 17.3%; Pred. No. 0.0067;  
Matches 335; Conservative 288; Mismatches 655; Indels 664; Gaps 90;

```

Qy 54 YKPPSSAKKVA-----NKDVASPLKEGLRL-SKELG-----LDEGSGVQLQCTVQLQ- 102
Db 862 YDHIQFVRAELMQLMKTLLRNPDASISHVAYRVLGKSSNRKMLKSKQKLVHYVTEVQG 921
Qy 103 ---EDYGRTRDSVKTVLQDEROSQALILKIID--YYEERTCILECVLHLTYFODER 155
Db 922 PSITVERSDCKAGIQLEMEKAIETALDCLKSANTEPYRRQAMEVIRCFVAMMSLEDNK 981
Qy 156 HP-YR-VEYADCVDKLEKELVSKYRQGFEEELYT--EAPYETNGNLTERRQVSRWFVQCL 212
Db 982 HALYQLAHNPFTEKIPNVIIISHR-----YKAQTPARKTEQALTGAFMSA-VIKDL 1034
Qy 213 REGSM-LLEIIFLYANFEWA-----PSDLVLVTWKPKQSGSGRQTNR 255
Db 1035 RPSALPFWASLIRHYTVAVAAQCGPFLPCYGVQSPS-----TAFHSEBNSK--- 1085
Qy 256 HLYDETMDPFVDRIGYFSALILVEGMDIESLHKCALDRRELHQFADGILICQMDICML 315
Db 1086 ---GMDP-----LVLDALAI--CMAYEEREL--CKIGEVALA 1116
Qy 316 TFGDIPHHAPVLLAMALLRHTLNPEETSSVVRKIGGTALQNLNFOYTRLLQSLASGND 375
Db 1117 VIFPV-----ASIIISKERACQLPLFSYIVERL----- 1145
Qy 376 CTTSACMCYV-----GLLS--FVLTSLEI--HTLGNQODIIDPACVLDPSLPEL 423
Db 1146 ---CACCYEQAMVAKLGAVSIIKFLMERLPLTWVLOQOFTKALLVMDLT--- 1195
Qy 424 FVGTEPTSGIILDSVCGMFPHLISPLQILRALVSGSKTAKVVSFLDKMSFYNELYK 483
Db 1196 ---GEVSNGAVMAAKTTLLEQLMRCATPLDEBERAEIVAAQEK-SFH----- 1239
Qy 484 HKRHDIYSH--EDGTLRRQTPKLLPLFGQTNLRI-----PGTVQVWLDRAIYVRW 536
Db 1240 HVTHTDLVREVTSPNSTVRKQAMSLQVLAQVTKSVIVINEPKHEVLAQDMVPPKGLLRH 1299
Qy 537 EYSSYSM-----TLFCEIEMLVHVSADVIQCOR-----VKPI 572
Db 1300 QPANAQIGMEGNTFCITLQPLRFLTMDLVNVEKHVFTELLNCEAEDSLTLPCYKSL 1359
Qy 573 IDLVHVKVISTDLSIADCLPRTSR--IYMLQRLTVISPPVDVIASCVNCLTVLAARN 629
Db 1360 PSIVPLRLIALMLAACNVLPQGRREKIIALFALNSTNSSELQAGACRKEFL-----E 1414
Qy 630 PAKVMTDLRHTGGLPYPAHVSSLSQMSISAGNAGYGLMLNBSQPOEYGTIAFLR 689
Db 1415 GATIEVDQHTHMRPL-----WMLGVRSLTN-----VVRRLT 1449
Qy 690 LITLVYQGLSTQSGCLVPCVNFLEKMLPSYHKRPNYNSHGVREQIGCLILEI----- 744
Db 1450 SVTRLFPNSFNDK-----PCDQWQHARKM-----MEVVVITHK 1483
Qy 745 ---HALINLCHETDLHSHSTPSLQF--LCISLAYTEAGQVIVINIGIVDTIMV 795
Db 1484 GGORSDNENISBCGRCP-----SPFCQFEMKICS-----AIIINLPHL----- 1523
Qy 796 MAQOPRDBAGGGQGLIKTYKLAFSYNNVIRLKPSPNVVSPLEQALSQGHAGNNL 855
Db 1524 IPAP-----QTLVKKPLLEVVMKTERAMLEAGSPREPPLKFLTRHPSQVEL 1572
Qy 856 IAVLAKYIYKHDPALPRLAIQLKRLATVAPMSVYACLANDDAARIDA-----FLTR 908
Db 1573 FMEKATL--NPPQSRMMSFLKH-----KQARPLRDVLAANPNRRIYL 1614
Qy 909 L-----QSKIEDMRIRKVMI--LEFTVAVEV----- 932

```

```

Db      1615 LLPGAGGTAVRGPSTSTWRLDLOFQAIIKIIISIVKDDSWLASGHSVLSQRLRWVSE 1674
Qy      933 -----OPGLIE-LFLALVYKDGSDGSKESRSL-----GMSCLHAUL-E 968
Db      1675 NFOERHREKMAATWTKERPKLAVCLVYCKKNYDIELLFOLLFAFGRFLCNMTFLIKE 1734
Qy      969 LIDSGOQRWYCCPRLLHRAIAFLAALMODRRDSAMLVLRTPKWENLTSPLF----- 1022
Db      1735 YMEETIPKQY---SIAQKALFFRVVDFFDPPFGDEL-----NAKVLQIHLNPAFLYSREK 1787
Qy      1023 -----GLSPSETSEPSILETCALIMKIICLEIYVYVXGSLDQSLKDTLKKEFSIEKRPAY 1078
Db      1788 GEGEQLLGPBPNEGPNESITSVFTKYLDPE-----KQADMLDLRLIYLQ-----Y 1835
Qy      1079 WSGYVYKSLAVHVALETGSSCTSLLEYQMLVSAMRLIITA-----TTADIMH-LTD 1129
Db      1836 ATLVVEHAPHHIDNNKRNRSKTL--RLIMTFAMPCLSKACVDPAKYSGHLLAHITAK 1893
Qy      1130 SVVRQQLFLVDLD-----TKAL-----LVPAVNCILRSGSMCTLLILLRQWK 1175
Db      1894 FAIHKKIVLQVHSLIKAHAMEARAIVRQAMAILTPPAVPAKEDGHO-----MLTMT 1946
Qy      1176 RELGSVDE--ILGPLETEILEGLVQ-----ADQOLMEKTKAKYFSAFIT----- 1216
Db      1947 RKI-IVEBGHTVPOLVHILHLVQHFKYVYRVHNLVQHMVSAMQRLGFTSVTIEQRRL 2005
Qy      1217 -----VLQMKEMKVSQDIPOYSQVLVNCETLQEBVIALPDQTHSLALASATEDK--- 1266
Db      2006 AVDLSEVVIKWEIQRIDKQOPSDMDPN---SSGCVNVSSSIKRGSLVSDAQKVKFR 2062
Qy      1267 -----DMETDDCSRSHRQORQCVCLGHLAKELCEVND----- 1303
Db      2063 TATGAISAVFGRSOSLPASDSLAKPDKQHTDTVNVN-LIRVAQVANDNTVTAGSPBEV 2121
Qy      1304 -----GDSWLQVTRRLPILFTLLTTEVLSLRMQ--NIHTEATILHLLTLTA 1348
Db      2122 LSRRCVNLKLTALRDMWPKSELKQWDFDKLMTYEQNQVYVNGICGLEVLSTLLVL 2181
Qy      1349 RT-----QOGATVAVAGITQISCLPILLSVYQLSTNGTAQTPSASRKSIDLAPSWP 1398
Db      2182 QSPALLSFPKLPORGIAACMTGNTK----- 2207
Qy      1399 GVVRLSWSLQOLKTLAYNPLPEALDFVGNHGERTLOCIANAVRTVOGLACLEADHTVG 1458
Db      2208 -VLRAVHSLRLKSI--FTEPSTSSVASKYER--LECLYA-----AVG 2247
Qy      1459 FILQ--LSNFMKEWHFHLPOLMRDIQV-----NIAYLCOACTSLHS--RMLQHYLON 1508
Db      2248 KVIYEGLTNYEKATYANPSQLFTLMIILKSACSNPSTIDRLISVPMASLQMKVREHLN- 2306
Qy      1509 KNGDGLPSAVAVQVRPPSASAPBSSKQPAADTEASEQOALHTVOYGLLKILSKTLAA 1568
Db      2307 -----FOAASGS-----TEA-----TSGSELVMLSLBLVK 2332
Qy      1569 LRHPTPVCQILLDLSLAEVNFALSFPTPTPDSVAPSGFGLLTVANVALMMLCEL 1628
Db      2333 TR-----LAVVSMEMRK-NFIQAIL--TSLIKESPDAKILRAVVKI-----VEW 2374
Qy      1629 DKKEPLOTAVGLSTQAGSTRLL--KSLIMFTMENCFTLLISQAWRYLRAVHPDRDOR 1686
Db      2375 VKANSFMA-----ANQPTLREKSLILLVKN-----NTYI-----EKR 2406
Qy      1687 MKQELSELSTLLSSLSRYFR 1708
Db      2407 FPEDL--ELNAQFLDLVNVYVR 2426

```

RESULT 35  
 US-10-424-599-256710  
 ; Sequence 256710, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:

```

; APPLICANT: La Rosa Thomas J
; APPLICANT: Kowalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 256710
; LENGTH: 2627
; TYPE: PR
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(2627)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: clone ID: PAT_MRT3847_73832C.1.pdp
US-10-424-599-256710

```

Query Match 1.7%, Score 156; DB 15; length 2627;

Best Local Similarity 17.0%; Pred. No. 0.0042;

Matches 326; Conservative 266; Mismatches 645; Indels 680; Gaps 83;

```

Qy      44 WRLLLEGLSTYKPPSPSAEKYKA-----NKDVASPLKELGLRISKPLGLDEQ 92
Db      392 WAPFESPPFKSSASSILRFIREILOENLCILHFRNVISAMNDL-----MEISSE 444
Qy      93 SVQLOCLQEDVGTDSVTVLODERQOALIKIADVYEEBRTCILRCVHLITVYFQ 152
Db      445 VILHRSFCR---KMQDLKQNSDFVDGTGRTVALMLGAVSCT--AHMCTVGANPSMLAEV 499
Qy      153 DERHPYREYADCVYDKLEKE---LVSKYRQOFELKYTEAPWETHGNLMTERQVSRWF 208
Db      500 DAVONPLVKSDCIGDMSKKAWSIIGALASFRLVSN-----SNGADETGK-----F 549
Qy      209 VOCIREQSMLEIIFLYIAYFEMAPSDLVLTQKFKEGQFGSROTNRLVDETMDFYDR 268
Db      550 LSLAKRYKSPQVLFVAVAGYLE-----FKGSLLEDAVYRIYHPELEKTKADA 597
Qy      269 IGVFSALIVGMDIESLHKCALDREHQAQDDGLICQMDCLMTFGDI PHHAPVL 328
Db      598 VATS-----DNLHSDKEXRISTIKLCHYKP--L 626
Qy      329 AMALLRHTLNEBETSV---VRKIGTAIQ--LNVFOYTRILQSLASGNDCTTSTACM 383
Db      627 GW-----ENSSVDQPAKRRKTEVSPTLNV-----ECTENNALL 660
Qy      384 CVYGLSFLVLSLHATLGNQODIIDTACEVLADPSLPELFWGTEPTSGLGIIIDSYCGM 443
Db      661 L--LSTETTPPISISSRSIQLFISKIOMELASGRIPNVY-----VPLVINGLFGI 709
Qy      444 -----FPHLSPLQLLALVSGKSTAKKVSFL---DKMSVYNLYKHKHPDVASHEDGT 496
Db      710 LNNRFSYLMNVLELCIAVLIS-----LHFLVMSVLAVALYERCO-----T 749
Qy      497 LMRQTPGLVPLAGGQTNLRIPOGTGCVMDRDAVYLVWMEYSYSWTLFTCEI--EMILLH 555
Db      750 IF--DTSNLH-----GSVNGALFDQAGLV-----DCKRLFYHNSDSGPS 789
Qy      556 VVSTADVIQHCQVKKPIID-----LVHNVISFD-----LSTADCLPITSRYMLLQRLTT 606
Db      790 VTIALLLQALQKIPVIEPRSRQFIPLEKFLGYPDLVSVGLFDSHACKKEMKALIKE 849
Qy      575 -----LVHNVISFD-----LSTADCLPITSRYMLLQRLTT 606
Db      850 WNLILKLMKPKSPYCGQFLKDVLOHRLLENDTEIQMRVLDCLLEBLT-----TW 899
Qy      607 VISPVDVIASC-----VNCLTVLAARNPAKWTDRHRTGFLPFVA--- 647
Db      900 SLGRESIIEBCHAYLVPLVIRLLMRVVRKTLKGLAKRKXASI---CHRKSILSFIINGLD 956

```

648 -----HPVSSLSOMISAEGNAGGYNNLNNSEOPQGEYGVIAFLITTLV 695  
 957 VVELPLFALLIKPQIYK-----TDGPALFMTSDKVSIDEPADALLEFTIDN 1008  
 696 KQGLSTOSQGLVPCVMPFLKEMLPYHAKWRYNHGREQ-----IGCILEL- 743  
 1009 IANLSMKKRYG-----LHVIEDII-----GVPELHIREPFLDLVGCVALLLES 1053  
 744 -----IHALNLCHETDLSHSPSLQFLCISLATEAGQVINIMGIVDTIDWMAAQ 799  
 1054 CTSSTHAAVNLGL-PSDQHN-----CSTSSNSLGE-----DSVPTNQ 1088  
 800 PRSDAEGQ--GQGLLIKTVLAFSYTNVNRILKPSNVVSPLEQALSQHAGNNLIA 857  
 1089 TQJNTLNQKDMBSLCKIITSL-----ATVAPMSVYAC-LGNDAAAIRDAFLT 1111  
 858 VLAKYIYKHDPALPRLAIQLIKRL-----ATVAPMSVYAC-LGNDAAAIRDAFLT 907  
 1112 VLANKYEDHEFSSDLMDREFSAVKPLVDKFOEASSEKPSLSLSCFLMSANNKLVALLY 1171  
 908 RLQSKIEMRIKVMLEFLYVAVERQPOLIELFLNLKFKDSDGSKERSLGMSCIAVL 967  
 1172 RKESLVPD-----IFSLISVNSASE-----AVIYCVL 1198  
 968 ELIDSQOQDRYWCPELHRAAIAFLHAILODRSDSAMLVLRTPKFWENLTSPLFGT--- 1024  
 1199 KFEV-----LISDNFENDBNSAQVLLSNIKVLMDSGCCIFGSNA 1242  
 1025 -----LSPSETSEPSILETCALMKIICLEIYVYVKGSLDOSLKDPLKFSIEKRPAYW 1079  
 1243 IKRKLIKPSGERT-----VIRILEPLPKYISEALIKOFDILLF-LINKTONS 1290  
 1080 SGVYKSLAV--HVAETEGSSCTSLLEYQMVASAMMLIITAHADIMHLDVSVRQLF 1137  
 1291 DRYEALQVITQITPILHGSTA-----XISAVSPLTISAB-----LDMRLRICDL 1337  
 1138 LDVLDTGALLLVPAVNCRLGSMKCTLLILRLQWK--RELGSVDEILGPLETEILE-- 1193  
 1338 LDALVADASALLSVAK-----LRLQWATSTLGMLDH-----DALINAY 1376  
 1194 GVLQADQOLMEKTAQKVPASATVYLOMKEMKVSIDIPQVSQVLAVNC--ETLOEVIYALF- 1250  
 1377 GIIMPD-----FFRSYOV-----EHALLILSHCVADMSSEETTFWS 1413  
 1251 -----DQTRHSIAL--GSATEDKDSMETDSCRSRHRDQPGVCLGHLAKLCEV 1300  
 1414 AYSSLSLVSFVDSAHILCOBGNSEBQLSVMKNTDSCWTSCIQTRAKKFLKHMADAM-- 1470  
 1301 DEDGD-----SMLQVTRRLPI-LPTL--LTTLEVSLRMKQNIHTEATLHLLTLARTQ 1352  
 1471 --DQSLAVIKGMIKLHQMWLKLPEVSNLKSJLWLCNDEGVNFPD----- 1514  
 1353 GATAVAGAGITQOSICPLSLSYQUSTNGTAQTPASRSLSLAPSV---PGVYLSLSLM 1408  
 1515 -----NITDSVY-----KRRVYALSMFRNVIVSNKFSBIT 1545  
 1409 EQLKTLRYNPLPEALDFVGVHOERTLOCLNAVTVQS-----LACL-----EBA 1453  
 1546 EKVPKRLFPNMLYDEKBEKAHEMKRA--CIETIASVSGQMKSYVALLICFPAGASRP 1603  
 1454 DHTVGFILQLSNFMKEWHFLPOLMRDIQVNIQVY-----CQAC--TSLIH 1497  
 1604 DKQKFLIRLCSILDKFHFSEVPNKEPKESLGGVSDMDITDTVNKEIQCLVAVLPK 1663  
 1498 SRKMLQHLQKN-----GDGLPSAVAVQRPSPASASAPSSSKQAPAD 1542  
 1664 IQKLNSDSERVNNISLAALKLKLPGDVMDIYLPITVIRISNPLKSHLESIRDEARS 1723  
 1543 TEAS--EQOALHTVOYGLIKLISKTALRHFTPDVQCILDDQSLDLAEVYFLAALFTY 1600  
 1724 ALATCMLGLEYQF--ILKVLQSTLR--RGYELHVLGYTL-----NFIISKCLSS 1771

1601 PTFDEVPAPSGFTLATVAVNALMGLDCKKEPLTQAVGLSTQAEGRITKSL--LMFT 1658  
 1772 PV-AGKIDYCLELDSLVIE--NDILGDVAQKE-----VENIASKMKETRRKSPESLKV 1824  
 1659 MENCFFYLLIQAMRYLRDPAVHPBDKQRMKQELSESLTSLSSIRYFRGAPSSPA 1715  
 1825 AONVTFK--SYVALKLAPVTAH-----LKKHITPYNKGLIENMLQHIANGIESNPS 1873  
 RESULT 36  
 US-10-369-493-2224  
 ; Sequence 2224, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; PRIOR FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 2224  
 ; LENGTH: 1557  
 ; TYPE: PRT  
 ; ORGANISM: Schizosaccharomyces pombe  
 ; US-10-369-493-2224  
 Query March 1.7%; Score 155; DB 14; Length 1557;  
 Best Local Similarity 16.2%; Pred. No. 0.0021;  
 Matches 303; Conservative 281; Mismatches 590; Indels 696; Gaps 83;  
 112 VKTVLQDERSQALLIKIADYYEERTCIDLRLHLLTYFQDERHPYRVEADCVDLK 171  
 2 IKLVDAEKGNM-----VNDASFERTWL-----WIPDKESFVKAM-----IVE 42  
 172 ELVSKTRQPE-----ELYTEAPTWETHGNLMTEROVSRRFVOCLEBQML-- 218  
 43 DLEKQFVKLERGSESEITVDGPAEKVNPPEKPMVMDMA-----LTCINEPSVNN 94  
 219 -----LEIFLYAVFEMAPSDLVL-----TRMKF----- 244  
 95 LTQRYEKDLIYYSGLFLVAVNPYCHLPYGDVVRKYQSKQKREKPHIFGTADAAYS 154  
 245 -----EOGFGSRQTRNHLVDETMDPFVDRIGYFSAALLVBSGMDIESLHKA 290  
 155 LBRRIHQSLVTVGESAGKETTKKVIQ-----YLSVNDASTSDQGLEKI 203  
 291 LDBRRELHQPAQDGLICQDMDC-----LMTFSDIPHHAVLALMALR--HTLNPEFS 343  
 204 LETNPVLEAFGNQYARNNSRFGKFIREFSNNGSIVGANIDWVLEKSRVYHPSNE 263  
 344 -----SVVRKIGTAIQ-----LNVQYVTRLLQSLASGANCCTTSTACGVYGL 389  
 264 RYHNVFQILRGADGSLBESLFLDRYVDHYSYKNGIKHI--NGVDDGKFPQLC-FGLR 320  
 390 SFVLTSLIHTL-----GNODIIDTACEVLAD-----PSLPFLMGTETPSG 432  
 321 TLGFNDNEHISLPLIASILHGN-----IEVASDRGQARFSLNQ----- 362  
 433 LGIILDSVCMFPHILS-PLQLLRALVSGSKSTAKVYISFLDK-----MSFINELY 482  
 363 -----IQOLC-----HLLEIPVDGFVNAALHFKSRYAGREWIPTARTBOVHTTQSLKGLY 414  
 483 KHKPHVDIYSHEDGTLWRQTPKLLYPLGGQTNLRIPQGTGVQVMDRAVLVMEYSYS 542  
 415 ERNPAHLVVRKINTQMTYSSSEH-----DGFIG--VLDIAGEFI--FTFNS 455



```

Qy 543 WT-----LFTCEIEMWLVHSTADVYIOHCORKEPIIDLVHKKYISDIL 584
Db 456 FEOICINFTNEKLOOFNHHYFVLEBOEYTERIEMDIDYGNIDQPIIDAEK--SEPI 513
Qy 585 SI-----ADCLLP-----TSRIYMLQRLTTVISPP----- 611
Db 514 GIFSCLEDCEWPMATDPTFEKHLHLFKGSDIYRPFKFSSEGVFLKHVADVEYDTKD 573
Qy 612 -----VDVASCVCNCLTVLAARNPAKWTDLRHGFLEPVVHPPVSSLSQMIASAGMANGG 666
Db 574 WLEKROPLNACLAALMFKSTN-----SHVSLFDDVYSNNSAG 611
Qy 667 YGNLNMSEBOQEX-GVTIAFLRLITLVKGOIGSTOSQGLVPCWMLKEMLPSYHKM 725
Db 612 RDN-----EKKGIFFTVSOHRROLSLMH-OLEAIOPH-FVRCI-----IPNNIKQ 657
Qy 726 RYN--SHQVREIOGL-ILELIH-----AIIMLCHETDHSHTPSLOFCISLAYTEAG 778
Db 658 PHNLDKSLVHLQKNCNGVLEGIRIAQOTGFPMKLFYTEFRARYGILSQSL---KRGYVAK 714
Qy 779 QTVINMGIGVDTIDMWMAAPRSDGAGOGGOLLITVYGLASVTNNVRLAP----- 833
Db 715 KATITI-----INELKL-----PSTVRLGETKVF 740
Qy 834 PSNVVSPLEQALSOHAGNNLIATV-----AKIY---HKHPALPRLAIQ 877
Db 741 KASVLSIED-----RRNALLRIVFNSFSARIRGFLTRRRRLRFRNRQAAI-LIQH 791
Qy 878 LKELATVAP-----MSVYACIG---NDAAIRDAFLTRLOSKIEDMRKIMILEFL 926
Db 792 NLROQLKLLKPPMWNFLHLKPLIGTQOTDEYLRKDALINNLOQLESTR---EVANEL 848
Qy 927 TVAVETQGLIEFLNLEVKOGSDGSKFSLGMSCLHAYVELDSOQDXYWCPPLLHR 986
Db 849 TITE-----RVQLTNDLOEQ----- 866
Qy 987 AAIAPFALMODRDSAMLVLRTPKFWENTSPFLGTLSPSESEPSILETCALIMKI 1046
Db 867 -ALHKEKDIIVERANSRVEVHERISLENOV---TIADEK----- 903
Qy 1047 ICLEIYVVKSLDOSLKDITLKFS-----IEKRFVMSGYVKSIAVHAETEGSSC 1098
Db 904 --YEFVLAEKOSIEEDLANKQTEISYLSDSLSTLEKLSIKOEQTISRYKELEKDYL 961
Qy 1099 TSLLEYQMLVASAMRL-----LIATTHADIMHLDTSVVRQLFD----- 1139
Db 962 NIMADYGHSSQHLNLEKANEKININLEKMLRLDDELILKORSYDTVOELREBNA 1021
Qy 1140 -----VLDTKALLVLPASVNCIRLGSMTCTLLIL--LRQKRELGSV---DELIGP 1167
Db 1022 SLKOOCRTYESQLASLVASKYSESELEKKEAEIVTFOKEITEYRDQHLKAPONEPETHN 1081
Qy 1188 LTELLEGLVQADQOLMEKTKAKVSAFTVLQMKEMKVSIDIPOYSQVLAVANCETLQOEVI 1247
Db 1082 INDVKSGLNSDENIY-STSTSTLSLKDVOELKSLHKEANQLSERKEISEMLEQSGIA 1140
Qy 1248 A-----LPDQTHSLALGATDKDSMETDCCSRHRDQDQGV 1286
Db 1141 TEEKLRKRNSELCDIIEALKTOIOOETEIIISLNA---DNLDLKDQNGVLEKNASDFI 1195
Qy 1287 CVLGL-----HLAKELCEVDEDDSMVQVTRRLPILPTLLITLLEVSILMKN 1333
Db 1196 DFOGKSRVEHKISDLNOLQERCKVG-----LNLKQK 1228
Qy 1334 LHFTATLHLTLTLARTOQATAVAGAGITQISILPLSLVYQLSTNGTAQTPSARSKSLD 1393
Db 1229 TENRSVQHTL-----DQNSPHSPFEKXSG 1254
Qy 1394 APSHPGVRL-----SMSLMEOQLKTLRYNPLPEALDVGVAHQEFTLQCL----- 1438
Db 1255 DP-----LKRIDGANDDRKIDNKLKLT-----ISKSLDALQTLVEBELSNLVSLSKDSLFT 1305
Qy 1439 -----NAVATVQ-SLACLEADHTVGFILQLSNFMKEMHFLHPQLMRDIOVNLGYLCQ 1490

```

```

Db 1306 DISGHINSRKLEKGLSTSELKER-----LMSNSDRP-----SPDIFKQDQ----- 1349
Qy 1491 ACTSLHSRKMLQHYLONKNGD---GLPSAVAQVQRP-----PSAASAPRS 1534
Db 1350 ---AINMSRKL-----LSNPNSDAQSLISSLOKTLVPESNMEPTGLKPLSPKISNLPSS 1402
Qy 1535 SSRQPADTEASEQALHTVQYGLKTLSTKLALRHF-----TPDYQQLLDDQ-----S 1584
Db 1403 S--QPSPSKRS-----GKHEALIRNPDONSSIPD--PFIWQNRNSVLO 1442
Qy 1585 LDLEAVNPLFALSTPTPDPSEVAPSFCTLIATV-----NVALNMLGELDKKE--P 1634
Db 1443 TEPEKILKLEAKRKSGILNDKDLKSTSELSIGLKKNEBELKULTSNLSDDGMALDPAP 1502
Qy 1635 LTOAVGLSTQ 1644
Db 1503 LLEDVPMNTR 1512

```

## RESULT 37

```

US-10-424-599-256711
; Sequence 256711, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(513223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 256711
; LENGTH: 2834
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_73833C.1.pcp
US-10-424-599-256711

```

Query Match 1.7%; Score 155; DB 15; Length 2834;  
 Best Local Similarity 17.5%; Pred. No. 0.0059;  
 Matches 346; Conservative 269; Mismatches 660; Indels 700; Gaps 89;

```

Qy 2 IRKSKITSVLSFCRS--SRBLMTILGR---SALRELSQIEFABINKHWRRLLEGISYKYP 56
Db 523 IFKASSSILRFTEILQENLCILHPRRNVISANNDMEISEBVIHLR----- 572
Qy 57 PSPSAEKVKANKDVASPELGLRISKFL-GLDEQGSVOLLOCYOE---DYGRTRDSV 112
Db 573 ---SFCERMQDK-----QNSDFVDGTSEBAPLTRICRLQEIICMGKXKINDI 618
Qy 113 ---KTVIODEKQNALIKIADYYEERTCILRCVCLHLITTYFOEBRHPYRVBADCVDKL 169
Db 619 AHADVLQIDBGVALALMGAVSCY--AHMCTVGANPSIMELVAVDNFLTVKSDCTGDM 676
Qy 170 EKE-----LVSKYRQOFEELYTEAPTWETHGNLMTERQVSWFVQCLREQSMLEIIFLY 225
Db 677 SKKAMESTIGALASFPRLYN-----SNHGADETGK-----FLSLAKRYKSSPOVLFAV 726
Qy 226 VAYFEMAPSDLLVLTNMFKEQGFSGRQTRNRLVDETFMDPFYDRIGYSALILVBMGDIES 285
Db 727 AGYLE-----FKHGSLLLEDAVYRIYHBELEKKTADAVATES----- 762
Qy 286 LHKCALDRLRHQPADGLICQMDCLMLTFGDIYRHARPLTLMALLRHTLNPEBTSV 345
Db 763 -----DNLHSDKEIRISTLKILCHYKP--LGM-----ENSSV 793
Qy 346 ---VRKIGTALIQ--LNVFYTLRLQLSLASGNDCTTSTACMCVYGLSLFTLSLELHT 400

```



```

Db 794 DQPAKCKRTEVSPILNV-----ECTENNALLL-----LISIEFTPISS 834
Qy 401 LGNODDIIDTACEVLAADSPLEPEMGTEPTSGIILDSVGM-----PPLISPLIQLLR 456
Db 835 SRSIOLFTSKIOEISAGRIENPV-----VPLVINGLFGILNRRSYLMNPVLEICIA 886
Qy 457 ALVSGKSTAKKVSFL-----DKMSFYNELYKHKPHVISHBOGTLMRQOTPKLYPLGQOT 513
Db 887 VLIS-----LHFLRWDSLVAYLERQ-----TIF--DIPSNLH----- 918
Qy 514 NLRIPOGVQVMDLDRAYLVYRMEXSYSSWTLFTCEI--EMLHVAVSTADVIQOCORVXI 572
Db 919 -----GSVNGALFPOPAGLV-----DCFKLFFVHASDSFSPVITLALLQALOKIPTV 966
Qy 573 ID----- 574
Db 967 IEPRSROPIPLFKELGYPDIVSVGLFDSHACKGEMKALIKEMWNLKLMKNPKSYCG 1026
Qy 575 -----LVHKVISTD-----LSIADCL-----LPITSRIYMLL-----ORLTT--VI 608
Db 1027 QELKOVLOHRLLENDTEIQKRVADCLIMWDYILPFYEHRLNLISSKNIREELITWSL 1086
Qy 609 SPPVDVVASC-----VNCITVLAARNPAKYWTDLRHTGFPFVA----- 647
Db 1087 SRESEIIIEECHRAVYLPVIRILMPRVKRLKGLASRKKASI---CHRSKISLFIAGLDIV 1143
Qy 648 -----HPVSLSQMISAGENNAGGYGNILMNSQPOGEYVITAFILRTITLVKG 697
Db 1144 ELPLFALLIKPLQIVKK-----TDGPANLFWTSDKVSIDEQADALLEVEFTLDNIA 1195
Qy 698 QLGSTOSOGIVPCVWFVKEMLPYHKRYNSHGREO-----IGCLILEL----- 743
Db 1196 NLSMKKKYGF-----LHVIEDII-----GPFDELHRRFLDLVGCVALLBECT 1240
Qy 744 -----HAIITLCHETDLHSHTPSLOPLCISLAYTEAGQTVINIGIVDTIDVMAAQR 801
Db 1241 SSLHNLNGL--PSDOHN-----CSTSNISLGE-----DSVPTQOTQ 1275
Qy 802 SDGAEQO--GOGQOLIKTVKLAFTVNNVIRLKPSNVVSPLEQALSGHAGNNLAVL 859
Db 1276 INGLINQJAKMRSLCKITISL-----VL 1298
Qy 860 AKYIYHKDPAIPLAIQLRL-----ATVAPWSVYAC--LGNDAAIRDAFLTRL 909
Db 1299 NKYEHEPSSDLMDFSAVXPVLDKFOZASSEKPSLISCLFASANNKVALLYRK 1358
Qy 910 QSKIEDAKIKMILEFLTVAVETQGLIELFLNLEYDQSGSKSEFSLGMSCHAVLEL 965
Db 1359 ESLVPD-----IFSISVNSASE-----AVIYCVLKF 1385
Qy 970 IDSQOQDRYWCPRLLHRAIAFLHALMODRDSAMLVLRTPKFMENLTSPLFGT----- 1024
Db 1386 VEN-----LISLDERPDEBNSAQRVLLSNIKYLMMSMCLFSSDNAIK 1429
Qy 1025 ---ISPSEISEPILTCALIMKIICLEIYVYVXGSLDQSLKDTLKKFSIEKFAVWSG 1081
Db 1430 RKLKSPGET-----VIRILEPLPKYISEABELAQFVILLLF--LENKTONSDV 1477
Qy 1082 YKSLAV--HYAETEGSSCTSLLEYOMLVGAMRMLLIATTHADIMHITSVARQFLD 1139
Db 1478 RVEALQVQNIITPILHGSTA-----KILSAVSPLYISAE-----LDMRLNICDLID 1524
Qy 1140 VLDTGKALLLVASVNCIRLGSMTLLILLRQMK--RELGSVEDEILGPTLEIL--GV 1195
Db 1525 ALVASDLSLSVAK-----ILRQLAMTSTIGMDH-----DALINAYGI 1563
Qy 1196 LQADQOLMEKTKAVFSAFITVLQMKEMKVSIDIPQSOVLNVC--ETLOEVIALF--- 1250
Db 1564 INTD-----FFRSVOV-----EHALILISHCVHMSSSETTFMFSAY 1600
Qy 1251 -----DQTHSLAL--GSATEDDSMETDCCSRRRDQDGCYVGLHLAKELCEVDS 1302
Db 1601 SSLSLFVDFSAHILCOBGNSEOLSVKMNIDSQWTKSCIOHTAKKFLIKHADM----- 1655

```

```

Qy 1303 DGD-----SWLOTRRLPT--LPTL--LITLSEVSLMKONILHFEATVILLITLARTQGA 1354
Db 1656 DGSLSVIKGIWIKLKHGXVNLKPEVSNLKSIMVLCNEDEGVNFFD----- 1699
Qy 1355 TAVAGAITQSICLPLISVYQLSTNGTAQFPASRSKSLDAPSW-----PGYRLSMSIMEQ 1410
Db 1700 -----NITDSVI-----RKRKVALSMFRNVIYVNNKSEFITEK 1732
Qy 1411 LKTLRYNPLPEALDFGVHQERTLOCLANAVTOS-----LACL-----EBADH 1455
Db 1733 VFKLFFNMLYDEKEGAEMKKA--CIETIASVSGQMKRSYALLIRFCWGSRSRDPK 1790
Qy 1456 TNGILOSMPKEMHNLQILMRDIOVNLGYL-----COAC--TSLHSR 1499
Db 1791 QKLFIRLISILDKFHFSEVPNHNKPEPSIGGSDMDITDPTVKNKEIOTCLYKVVLPKIQ 1850
Qy 1500 KMLQHYLQNRX-----GDGLPSAAGVQRPASASAPSSKOPADTE 1544
Db 1851 KLANSDEKYNVNIISLAALQLLPLRQVDVLYLPTVHRISNPLKSHLESIDEARSAL 1910
Qy 1545 AS--EQALHTVOYGLKILSKTILALRHFTPDVCOILLDQSLDLAEYNFLPALSTFTPT 1602
Db 1911 ATCLKELGLELYQF--ILKVLOSTLR--RGYELHVLGYTL-----NFLSKCLASSPV 1958
Qy 1603 FDESVAPSGTILATVAVNALMTEGDKKEPILTOAVGLSTOABGRTILKSL--LMTME 1660
Db 1959 -AGKIDYCLEBLLSVIE--NDIIGDVAEQCE--VEKIASKMKETRRKRSFESLKVIAQ 2011
Qy 1661 NCFYLISQARYLRPAVHPRDKQMKQELSELSTLSLSRYPRGAPSSPA 1715
Db 2012 NMTPE--SYALKILAPVTAH-----LKGHITVNVGKLEMLQIATGIESNPS 2058

```

```

RESULT 38
US-10-723-860-2568
; Sequence 2568, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882,0193,NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2568
; LENGTH: 3830
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-2568

```

```

Query Match 1.7%; Score 153; DB 17; Length 3830;
Best Local Similarity 17.2%; Pred.No.0.015;
Matches 331; Conservative 285; Mismatch 658; Indels 654; Gaps 88;

Qy 54 YKPPSSSAEKVVA-----NKDVASPLKEGLRI--SKFLG-----LDDEQSVQLQCYLO- 102
Db 862 YDHIOQVRAEMALMQLRTLRNPADSIASHAYVYLGFGGSSNKMRLKESQKLAHVYVEVQG 921
Qy 103 ---EDYGRTRDSVKTVLQDERGQALILIKIAD---YVVEERTCILRCVYHLITTYQDER 155
Db 922 PSITVEFSQCKASLQAPMEKALFETALDCKSANTEBYRQAWEVYKCFVLVAMGSLJDK 981
Qy 156 HP-YR--VEYADCVYDKLEKLVSKYRQOFELVYKT--EAPFWETHGNLMTERQVSRWVQCL 212
Db 982 HALYQLAHNPTEKTIIPNVIIISHR-----YKAQDTPARKTFEQLATGAFMSA-VIKOL 1034

```

213 REOSW-LLEIFLYAYFEMA-----PSDLLVLTKMKEQFGSGRQNR 255  
D 1035 RPSALPFAASIRHYTWAVAVAOQGPFLPCYQVSGQS-----TAMHSENSGK 1085  
Q 256 HLVDEMDPFVDRIQYFSAIILVEGMDIESLHKCALDRREIHOAOGLIQQMDICML 315  
D 1086 -----GMP-----LVILDAIAI-----CMAVEKEEL-----CKIGEVALA 1116  
Q 316 TFGDIPIHAPVLLAMALLRHTLNPEETSIVVRKIGCTAIQANVPOYLTRLLQSLASGND 375  
D 1117 VIFPV-----ASILSGKERACQLPESYIERL----- 1145  
Q 376 CTSTACWCYV-----GLLS--FVLSLEL--HTLGNODIDTACEVLADPSLEL 423  
D 1146 -----CACCYEQANVYALGVSISKIFMERLPLTWVQNOOTFLKALLPFMMDLT 1135  
Q 424 FWGTEPTSGIGIILDSVCGMPHILSPILLQRLALVSGSKTAKKVSFLDKMSFYNELYK 483  
D 1196 -----GEVSNGAVAMAKTLEQLLMRCATFLKDEBAEIEVAOER-SFH----- 1239  
Q 484 HKPHDIVSH--EDGLMRQTPKLYPLRGOTNARI-----PGTVGQVMLDDRAIYVRM 536  
D 1240 HVTDLVAREVTSPTSSTVRKQMGSLQVLAQVTKSVTVMPEHKEVLQDMVPPKGLLRH 1239  
Q 537 EYSYSW-----TLFTCEIEMLHVSTADVIQHQR-----VKPI 572  
D 1300 QPANAQIYMESGNTCTTLOPRLFTMDLNVVHKVFTYELNLCEAELSALTKLPCYSLS 1359  
Q 573 IDLVHNVISTDLSIADCLPITSR---LYMLQRLTVISPPVDIASCVNCLTVLAARN 629  
D 1360 PSLVPLRIALNALAANCYLPOSREKIIAALFKALNSTSELQSEAGCMKFL-----E 1414  
Q 630 PAKWTDLRHGFLPFVAPVSVSSQMSIAGGNAAGVGNLMSBOQGEVGTIARL 689  
D 1415 GATLEVQDITHMPEL-----MMGLDRSLTLN-----VVRMLT 1449  
Q 690 LITLVKQGLSTQGLVPCVMFVLEKMLPSYHKNRSHGVBEQIGCLILIELHALN 749  
D 1450 SVTLFPSPFNDK-----PCDOMQHRLKRM-----EVVV 1479  
Q 750 LCHETDLHSHTPSLQFLCISLAYTEAGQVIVINIGIVDTIDVMAAOPRSDGAEQ 809  
D 1480 ITHKGQSGDGNB-----MKICS-----AIIINFLH-----IPAP 1510  
Q 810 QGOLLKTVKLAFSVTNNVIRKLPSPNVSPLEQSLSHGAGNLLVLAITYHKDP 869  
D 1511 --QTLVVKLELVVMKTERAMLEAGSPREPILKFLTHPSQVLEFVMEATL-----NDP 1564  
Q 870 ALPRLAQLKRLATVAPMSVYACLGNDAAAIRDA-----FLTRL-----Q 910  
D 1565 QMSKMFNSFLKH-----KQARPLRDVLANPNRFTILLPGANQTRVPRGS 1610  
Q 911 SKIDMRKIKWI--LEPLTVAET----- 932  
D 1611 PSTSMRDLQFOAKIKIISIIYKNDSDWLASOHSIVSOLRRVWSENQERHKKENMAAT 1670  
Q 933 ---QPGLE-LFLNLEVQDSDGSEBSL-----GMSCLHVL-ELIDSOQOORRYCPR 982  
D 1671 NMXEPKLAAYCLINVCCKNVDIELFQLRAFTGRFCNNMTFLKEMVEBEIPIKXY--S 1727  
Q 983 LLHBAALFALAMODRDSAMLVTRKPKFMENTSLPLF-----GLSPSPETS 1032  
D 1728 IAOGRALFFRVDFDPNFBDEL-----KAKVLOHILNPAFLYSEKGEGBOLGPBNBEG 1783  
Q 1033 EPSILTECALIMKIIIEIYVVVKSLSQSLKDTLKKSIEKRFAYMSGYKSLAVHVAE 1092  
D 1784 DNPSITSVFTTKVLDP-----KQADMDSIRIILQ---YATLVENHAPHHHD 1831  
Q 1093 TEGSSCTSLBEXOMLVSAWRMLIYA-----TTADINH-LTDSVVRQQLFLVDIG 1143  
D 1832 NNKARNKSL--RRLMTFAMPCLLSKACVDPACKYSGHLLAHIIAKPAIHKKIYLQVPHS 1889  
Q 1144 -----TKAL-----LVAVASVNGLRIGSMKCTLLILLRQMKRLSGVDS--ILGP 1167

D 1890 LKKAHAEARAIQVQAMAILTPAVPAMEDQH-----MLTWTBTKI--IVERGHTVPQ 1941  
Q 1188 LTELLEGVLQ-----ADQMLEKTKAKYFSAFIT-----VLQMKEM 1223  
D 1942 LVHILHLIVQHFYKVVYVRHHLVQHWVSAMQRLGFTSVITIBORRLAVDLSEVVIKWELO 2001  
Q 1224 KVSIDIPOYSOLVNVCESTLOEVALPDPQTRHSIALGATEDK-----D 1267  
D 2002 RIKQOQPSDWDPR-----SSGEGVNSVSSSIKGLSVSADQVYKFRATATGISAIVFGRSQ 2058  
Q 1268 SMTEDCSRSRHRDQDGVCVLGHAKELCEVED----- 1303  
D 2059 SLPGADSLAKPIDKQHTDVTYVNF-LIRVACQVNDNTNTGASPEVLSRRCVNLKTAALR 2117  
Q 1304 GDSWLOVTRRLPILPTLLTLEVSLEKQKQ-NLHTEATLHLILTLART-----QQ 1352  
D 2118 PDMMPKSELKLOMEDKILMTVEQPNQVNYGNICTGLEVLSFLTLVLOS PALLSFKPLQR 2177  
Q 1353 GATAVAGAGITQISCLPLLSVYQSTNGTQTPASRKSIDAPSPGVYRLSMIGQL 1412  
D 2178 GIACMTCGNTK-----VLRVHSLSRML 2202  
Q 1413 KTLRYNLEPEALDEGVHQBERTLOCLNAVRTVQSLACLEADHTVGFILQ--LSNFMKEW 1470  
D 2203 SI--FPTPESTSSVASKYEE--LECLYA-----AVGKIVYEGLTNYEKAT 2243  
Q 1471 HFLPLQMRDIQV-----NLGYLCOACTSLHNS-RKMLQHYIQRNKGDLPSAVARV 1522  
D 2244 NANPSQLFGTLMILKSCASNPNYSYIDRLISVFMSRLQKMYREHLN----- 2288  
Q 1523 QRPSSAASAPSSSKQAPADPEASEQALHTVQGLKIITKTLAALRHFTPDVQCILLD 1582  
D 2289 ---POAASGS-----TEA-----TSGTSELVMSLELVKTR-----LAV 2319  
Q 1583 QSLDLAEYNLEFALSFTPTPFDSVAPSGTLLATVVALNMLGELDKKEPILTOAVGLS 1642  
D 2320 MSMMERK-NFIQAL-----TSLIEKSPDAKILRAVNI-----VEWYKNSPMA----- 2364  
Q 1643 TOAEGRTTL--KSLMTMENCPLYLLISQAMRYLRDPAVHRDQKQKQKQKQKQKQKQKQ 1700  
D 2365 --ANQTFLEKKSILVYM-----MTYI-----EKREPEDL--ELINAQFL 2400  
Q 1701 SLGRYFR 1708  
D 2401 DLVNYVYR 2408  
  
RESULT 39  
US-10-437-963-108860  
Sequence 108860, Application US/10437963  
Publication No. US2004012343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yinhua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 108860  
LENGTH: 2159  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_13073C.1.pcp

US-10-437-963-108860

Query Match 1.7%; Score 152.5; DB 16; Length 2159;

Beet Local Similarity 18.7%; Pred. No. 0.0062; Matches 328; Conservative 243; Mismatches 574; Indels 609; Gaps 82;

227 AYFENAP--SDLLVLTQFKKEG-----FGSRQTRHLVDETMPFV-----DRI 269  
 213 AIFEVSSGSLDDHCKMIFVTEGVVPTLMDLTKRSH-QDKVGEFTGALRNLCGKD 271  
 270 GYFSAALLVEGMDISLHKCALDRRELHQPADGLICQDMCLMTFGD-IPHHAPVL 328  
 272 GYMRANLAGEVEIIT--GLISSKTTSSQSNABSLAR--LVSAGDSIPKIDAGA 324  
 329 AMALRHILNPE-----ETSSVVRK--IGGAIQIANFOYITRLQ 367  
 325 VKALR-LNKRNDISVRESADALEALSSKSIKKAVDAGGIPV-----LIG 373  
 368 SLASGNDCTTSTACMVCYGLSPVLTSLHTLGNODIIDTACEVLADPSLPBLFWGT 427  
 374 AVVAPEKQMGDTG-----HSLQSHAVHALSN-----IC----- 403  
 428 EPTSLGIILDSVCMFPHLSPILLQILRALVSGSKTKKYSPFLDKSGFYNELKKEH 487  
 404 GGTVSLILLYIGELC-QVPSPPVPLADII GAL--AYTLWFSGTGKSF----- 448  
 488 DVISHEDGTMRROTPKLLYPLGGQTNLRIPQGTVGVMDDRAVLVWEXSYSMTLFT 547  
 449 DPIEL-----NILI-----VLK-----STQSNVLVD 471  
 548 CEIEMLHVSTADY--IOHCQVKKPIIDLVHKVISTDLSIADCLPITSRIYMLQRL 604  
 472 RILEMALSIGNACISGRILNSNAKVALVGLI----- 503  
 605 TTVISPPVDVASCNCLTVLAARPAKWTDLRTGFLPVAHPVSSLSQMSAEGMNA 664  
 504 -TMS--ADVQKNLVHALTSLCS--DGIGIMALGR-----EGTOLLISFLG-- 546  
 665 GGYGNILMNSRPOGEYGVITAFELRITLVKQSGTOSQGLVPCVMFVKEMLPYHK 724  
 547 -----LSRHOQ-EYAVSL--LAISDEVDDSKMTAAGGIPVLQILET----- 589  
 725 WRYNHGVREQICILIELHAILNL-CHEIDL----- 756  
 590 ---SQKAKE-----DAHILMWLCHSDDISACVESAGAVLALLMLKSGSPHQEA 639  
 757 -----HSHPTSLOFLCICSLATYEA-GQTYINIMSGVDTIIDMVAQAQRSD 803  
 640 SAKALKIIRSAADSTINQLRALISDLSSTKHAHITVIGHVLMASQORDLVONGAPANK 699  
 804 G-----AEGQGGQLIKTVLAFSVTNNVIRLKPSPNVVSPLEQALSQGH 851  
 700 GLRSLIDLESSNETEQATVADI--FSTRODICDIIGTBEIIOPCMKLITS-- 752  
 852 GNNLIATLAKTYHKHDPALFRLAIQLIKRLATVAPMSVVAACGNDAARDAFLTRIQS 911  
 753 GNOVIAT-----QSRALGALS--HSANMLKKRKSICIAEGVQTLIE 793  
 912 KIEEMRIK-----VMILEPLT--VANETOPGILFLFMLEVDGS--DGKESLSGMS 961  
 794 MSKSPSIDAAETTIAALNPLSDAHIAKALDGNIVALTIVLKEGLEKGISASRLCQ 853  
 962 CLH-AVLELIDSOQODRYMCPPLHRAAIAFLHALMODRDSA-----MLVLRTPK 1012  
 854 LLNGPLNEVIPDSQ-----CYFIH--ALIVCLSGINENNTNDPLNVLAMARXK 905  
 1013 FWEMLTSPLEGTSPSETSE-----PSLETICALIMKICLEIYVYVKSGLDSQ 1062  
 906 EGAFHSSSLWGAFLDVPESLEPLVACISVGLPPIQDKAIQITAILCQDOPSLIEHLNRS 965  
 1063 LKOTLKRSIEKRFAYMSGYKSLAVHYAETE-----GSSCTSLLEKQMVSAWMLLI 1116  
 966 -----OGCTIASLASRVISTMEIRIGSAIT-----LISMR----- 997

1117 IATTHADIMHLDTSVNRQLFLVDIOCTKALL--LVPAVNCILRGSKCTLLILRLQWK 1175  
 998 ---HS-----REHSIDVTEASGHUKNLSISIDMKODSAPTL--DIEWVK 1039  
 1176 ---RELGSVD-EILGLPTEILSGVLQADQOLMEKTKAKFSAFTYQLW-KEMKVS DIP 1229  
 1040 PYPENSLYNDKVOVG-----VSGSGKVLIEETALWLLSLICSHSSKULTVWD-- 1088  
 1230 QYSOLVINCETLOEVIYALPDOTRHSLSALGATBEDKOSMETDCSSRRHRODQVCVL 1289  
 1089 ---LGVETISDKLASV-----TTNQODSMLVQSPA----- 1116  
 1290 GHLAKELCEVDEGDSWGLQVTRPLITPLTLLEVSLSMKONHFTETATL-HILLTLA 1348  
 1117 -----IMRTISLASLSKDKIT-----DKTPAOSLASLSTGS 1151  
 1349 RTQOGATAVAGA-----GITQSICLPULSV--YOLSTNGT-----AQT 1385  
 1152 RSIQLATANSAGVWGTIAMIGIESWPNLVMAABEFKLDNPSKILRSLFELEDVRTS 1211  
 1386 SASRKS-----DAPSNPQYRSLMSIMEQLTKRYNPLPEALDPVGHQOERTLOC 1437  
 1212 ATARRSIPLDVLDLKEWPDROGAPLVALHLLTOLABGSESTNKV--AMAEAGV----- 1261  
 1438 LNAVRTVQSACLEADHTVGFILQLSNFMKEWHFHLPOLMRDIOVNLGYLCACTSLH 1497  
 1262 LDALTKLTSIPDSSTETITINILRIITYNPDLIYHSSISTSNQVAVARLSRSRN 1321  
 1498 SRKLOHYLONKQGDGLPSAVAQRVORPPSAPASAPSSKQAPADTEASEQAL----- 1551  
 1322 AARTQLNLPSEN-----IRDTEVAMQAIPLDMLBSGETEGQALGALIKL 1370  
 1552 -----HTVOGLAKIIS-KTLAARHFTPVQCIIIDQSLDLA----- 1589  
 1371 SSGNISKASALFVEGTTLESILKYILSFSSLSLEKDAQALCYILENSTIRSPASEC 1430  
 1590 YNPLFALFTPTFDESVAPSFGLTATVAVNALMGEIDKKEPTLQAVGLSTOAGTR 1649  
 1431 LQPLISLMTSGSTFVEBPANR-----ALNRLLDEVTNME-----IAATSEVD 1473  
 1650 TLKSLMFT--MENCFTYLLISQAMRYLRDPAVHPDRQKRMKOL----- 1691  
 1474 LTVSPVGVTHQJSEACIGALIKG-----KDRNCKLEWVKAGIIEHVLMDIL 1522  
 1692 ---SSELSTLSLSKRFRRGAPSSPA-----TGVLSPQCKSTSLSKASP 1734  
 1523 DVEPVSSSTIAELIRLITN--NSGIAKSSAAAMHSALQALVNLKPPQ--SLAALKLSP 1578  
 1735 ESO--EPLIQLVOA 1746  
 1579 -SQIIEPLISFLS 1591

RESULT 40  
 US-10-408-765A-214  
 ; Sequence 214, Application US/10408765A  
 ; Publication No. US20040101874A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ghosh, Soumitra S.  
 ; APPLICANT: Fahy, Eoin D.  
 ; APPLICANT: Zhang, Bing  
 ; APPLICANT: Gibson, Bradford W.  
 ; APPLICANT: Taylor, Steven W.  
 ; APPLICANT: Glenn, Gary M.  
 ; APPLICANT: Wainock, Dale R.  
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
 ; FILE REFERENCE: 660088.465  
 ; CURRENT APPLICATION NUMBER: US/10/408,765A  
 ; NUMBER OF SEQ ID NOS: 3077  
 ; SOFTWARE: FastSeq for Windows Version 4.0



